ERIC Document Resume | See Instruction Sheet | DATE OF RESUME | DATE RECEIVED

A. PUBLICATION DATA

1. ERIC NUMBER: ED002172
2. ORIGINATOR CONTROL NO(S).: N.A.
3. IDENTIFICATION NO.: N.A.
4. AUTHOR(S): Diederich, Paul B. and others
5. TITLE OF DOCUMENT ABSTRACTED: Factors in Judgments of Writing Ability
6. PUBLICATION TITLE: Research Bulletin
7. EDITOR(S): N.A.
8. PLACE OF ORIGIN
ORG. ST. Educational Testing Service
CITY, STATE Princeton, N. J.
TELE. EXT.
INST/LOC CODE
9. PUBLISHER OR SOURCE: Educational Testing Service Princeton, New Jersey

10. VOL.	11. PAGES	12. MO./YR.	13. TOTAL PP.
N.A.	N.A.	8/61	99

DO NOT WRITE IN SPACE BELOW

MICROFILM COPY AVAILABLE YES ☒ NO ☐

COPYRIGHT GRANTED YES ☒ NO ☐

COPYRIGHT RELEASE FILED YES ☒ NO ☐

TYPE OF REPORT
PROPOSAL ☐
IN-PROGRESS ☐
FINAL ☒
FOLLOW-UP ☐

OTHER ERIC REFERENCES:

ERIC SATELLITE CODE

B. INDEXING DATA

14. SUBJECT AREA
15. TYPE OF ENDEAVOR: experimental study
16. INDEXING TERMS (6-12)
college level
English composition
scoring writing ability (grading)
factor analysis
Educational Testing Service, Princeton, N.J.

17. ABSTRACT (MAX. 250 WORDS)

Unfavorable results in grading essays, particularly the College Entrance Examination Board General Composition Test, led to the study. It was expected that progress in objective evaluation of essays would be made, by means of a factor analysis of the judgments of a diverse group of competent readers in an unstructured situation. It was hoped that some qualities amenable to objective testing would evolve and that more precise observation of the remaining qualities could lead toward closer agreement in scoring essays.

About 600 short papers were secured from freshmen in three colleges requiring the College Entrance Examination Board Tests. Ten representatives known to be concerned with student writing, from each of five fields, participated as readers. The readers used their own judgments as to what constitutes writing ability, sorted the papers into piles according to merit, and wrote brief comments on likes or dislikes about the paper. Classification of comments and development of the data resulted in 55 categories of comments which were in turn reduced to five factors--ideas, form, flavor, mechanics, and wording. Factor analysis was applied.

It is concluded that the task of readers may be simplified by restricting attention to ideas, form, and flavor. Mechanics and wording can be measured more reliably and systematically by objective tests. Additional studies are needed to produce more precise observations of these factors to lead to progress toward agreement on common standards.

ERIC No.

18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION
20. DIVISION OR DEPT.
21. CONTRACT NO.
PROJECT NO.
22. TOTAL COST OF PROJECT
23. DURATION OR PROJECT

U.S. Department of Health, Education, and Welfare
Office of Education

ERIC Form 1

ED002172

RESEARCH BULLETIN

filed under The Arts TD

RB-61-15

Beyan

0030

BULLETIN

FACTORS IN JUDGMENTS OF WRITING ABILITY

Paul B. Diederich
John W. French
Sydell T. Carlton

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
August 1961

REPRINTED BY PERMISSION
U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN PRINTED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FACTORS IN JUDGMENTS OF WRITING ABILITY

Abstract

The purpose of this study was to serve as a stepping stone toward closer agreement among judges of student writing at the point of admission to college by revealing common causes of disagreement. It was expected and found that more than half the variability in grades of a large number of judges on the same set of papers was due to "error" (random variation) or the idiosyncratic preferences of individual readers. In the variability that was not random or idiosyncratic, it was expected and found that there was a substantial core of common agreement on the general merit of the papers and that a small number of "schools of thought" would account for most of the systematic differences in grading standards. Factor analysis of correlations among the grades of 53 distinguished readers, representing six different fields, on 300 papers written by college freshmen of widely varying ability revealed just five such "schools of thought," emphasizing:

I. Ideas: relevance, clarity, quantity, development, persuasiveness;

II. Form: organization and analysis;

III. Flavor: style, interest, sincerity;

IV. Mechanics: specific errors in grammar, punctuation, etc.;

V. Wording: choice and arrangement of words.

No standards or criteria for judging the papers were suggested to the readers. Instead, they were told to use "whatever hunches, intuitions, or preferences you normally use in deciding that one paper is better than another." They sorted the papers into nine piles in order of "general merit." The only restriction was that all nine piles must be used, with not less than 4% of the papers in any pile.

The five reader-factors or "schools of thought" were identified by a "blind" classification of 11,018 comments written on 3,557 papers that were graded high (7-8-9) or low (1-2-3) by the three highest and three lowest readers on each factor. The person who classified the comments did not know the standing of any reader on any factor. In addition to the reader factors, three College Board tests taken by these students formed a separate "test-factor" that had practically zero correlations with all reader-factors except Mechanics (.50) and Wording (.45).

It was not the purpose of this study to achieve a high degree of unanimity among the readers but to reveal the differences of opinion that prevail in uncontrolled grading--both in the academic community and in the educated public. To that end, the readers included college English teachers, social scientists, natural scientists, writers and editors, lawyers, and business executives. None the less, it was disturbing to find that 94% of the papers received either seven, eight, or nine of the nine possible grades; that no paper received less than five different grades; and that the median correlation between readers was .31. Readers in each field, however, agreed slightly better with the English teachers than with one another.

TABLE OF CONTENTS

Glossary for English Teachers 1

Background and Purpose 1

How the Papers Were Selected 3

How the Readers Were Chosen 8

How the Papers Were Graded 10

The Intercorrelations 13

The Factor Analysis 15

How the Comments Were Classified 18

Analysis of Comments by Factors 23

Interpretation of Factor I 28

Interpretation of Factor II 33

Interpretation of Factor III 36

Interpretation of Factor IV 38

Interpretation of Factor V 39

Interpretation of Factor VI, the Tests 41

Summary of Interpretations of Factors 41

Correlations with "Consensus Scores" 44

Effort to find "Differentiating Papers" 46

Low Correlations of Test Scores with Consensus 48

Correlations with Objectively Scorable Quantities 49

Leniency-Severity 50

Discussion and Conclusions 51

References 59

Acknowledgements 60

TABLE OF CONTENTS (Cont'd)

Appendix A Further Details of the Factor Analysis A1

Appendix B Table of Intercorrelations B1

Appendix C Theme Topics C1

Appendix D Readers D1

Appendix E Correlations of Readers and Tests with Consensus Scores E1

Appendix F Distribution of Grades for Each Reader F1

Appendix G Distribution of Grades for Each Student G1

Appendix H Sample Score-Sheet H1

Appendix I Review of John B. Carroll, "Vectors of Prose Style" I1

Appendix J Glossary for English Teachers J1

LIST OF TABLES

Table 1	How Many Readers Placed Every Thirtieth Paper in Each Pile	12
Table 2	Median Correlations	14
Table 3	How Seven English Teachers High on "Mechanics" Graded these Papers	17
Table 4	How Seven Readers High on "Ideas" Graded these Papers	17
Table 5	Analysis of Comments of Highest and Lowest Readers on Each Factor	24
Table 6	Correlations among Three Highest Readers on Each Factor	34
Table 7	The Fields of Readers Having Factor Loadings of .25 and Above	43
Table 8	Correlations with "Consensus Scores"	44
Table A-1	Unrotated Orthogonal Factor Loadings	A7
Table A-2	Oblique Factor Loadings	A8
Table A-3	Factor Intercorrelations	A9
Figure A-1	Plots of Factor Loadings Before and After Rotation	A10
Appendix B	Table of Intercorrelations	B1
Appendix E	Correlations of Readers and Tests with Consensus Scores	E1
Appendix F	Distribution of Grades for Each Reader	F1
Appendix G	Distribution of Grades for Each Student	G1

FACTORS IN JUDGMENTS OF WRITING ABILITY

Glossary for English teachers

It is hoped that this report will be read by English teachers, many of whom are unfamiliar with factor analysis. A brief non-technical explanation of the key terms and processes is therefore given in Appendix J. Readers who are uncertain of the meaning of such terms as "correlation" and "factor" should start with this appendix.

Background and purpose

The most comprehensive review of studies of the reliability of grading English compositions for writing ability appears in Huddleston (1954). Since this journal is readily available, it seems unnecessary to spend several pages in a review of the literature to demonstrate what everybody knows: that grading essays for writing ability, even under the most carefully controlled conditions, is extremely unreliable.

Since Huddleston's study, the College Entrance Examination Board spent six years and a grand total of $97,700 on the development and tryout of its General Composition Test, a two-hour essay test of writing ability. In this test, an effort was made to increase reliability by

1. presenting extensive "stimulus material" for background and content;
2. training the readers to grade on five carefully defined qualities (mechanics, style, organization, content, and reasoning) rather than giving a single grade;
3. grading each quality on a four-point scale:

    1- unprepared for college work
    2- needs remedial training
    3- needs the freshman composition course
    4- might be excused from the freshman composition course;

4. allowing borderline grades whenever a paper did not fall clearly within one of these four categories;
5. resolving differences in grades by using a third and even a fourth reader when the first two did not agree.

BLANK PAGE

Final figures for the inter-reader reliability of grading under this system were given by Swineford (1956), based on the last administration of this test in May, 1956. They ranged from .46 for reasoning to .66 for mechanics, and the reliability of the composite score (the sum of the scores on the five qualities) was .69.

This, however, is the reliability of the scoring, not the reliability of the examination, which has to be determined by correlations between scores received by the same students on two different examinations of this type. Such correlations were reported by Olsen (1956) for 226 students in six well-known private preparatory schools who took the General Composition Test as juniors in 1954 and again as seniors in 1955. The correlations were .25 for organization and reasoning, .26 for content, .37 for style, .46 for mechanics, and .44 for the composite score.

Their scores on the General Composition Test were also compared with ratings on these same qualities, using the same definitions and score-sheets, by the English teachers in these schools who had just taught them for a full year. Correlations between test scores and ratings on the five separate qualities ranged from .22 on content in grade 12 to .44 on style in grade 11, with most of the other figures lying between .30 and .40. Composite ratings (the sum of the five separate ratings) had lower correlations with composite GCT scores than with scores on two shorter College Board objective tests, as shown below:

Correlation of teachers' composite ratings with	Grade 11	Grade 12
General Composition Test composite scores	.52	.44
College Board English Composition Test (one hour)	.65	.61
College Board SAT Verbal (90 minutes)	.70	.65

It is also noteworthy that the correlation between SAT-Verbal scores from one year to the next was .88. Thus, the Scholastic Aptitude Test not only picked out the good and poor writers (in the judgment of these teachers) each year better than the essay, but it picked out almost the same students, while the essay tended to pick different ones.

At this point it was agreed that further progress in grading essays must wait upon a factor analysis of the judgments of a diverse group of competent readers in an unstructured situation, where each could grade as he liked. As the abstract indicates, it was believed that such an analysis would reveal "schools of thought" in grading essays that would account for most of the systematic differences in grading standards. It was also hoped that some of these "schools" would emphasize qualities amenable to objective testing, and that more precise observation of the remaining qualities could then lead toward closer agreement.

Funds were provided by the Carnegie Corporation; the cooperation of colleges and readers was secured; and the study outlined in the following pages was undertaken.

How the papers were selected

During the month of November, 1958, Educational Testing Service secured about 600 short papers written as homework on a choice of four different topics by a representative sample of liberal arts freshmen in Cornell, Middlebury, and the University of Pennsylvania. These three colleges were chosen because they enroll both men and women and require both the Scholastic Aptitude Test and the English Composition Test of the College Entrance Examination Board.

An initial inspection of the papers revealed that too few had been written on two of the four topics to secure a wide and even distribution of SAT Verbal scores. Hence these topics were excluded and papers were chosen on the remaining topics in such fashion that the two distributions of SAT Verbal scores were parallel to one another.

It is worthy of note that the two rejected topics dealt with concretely described incidents that were supposed to be close to the interest and experience of students: what to do about a popular student treasurer who was neglecting his duties, and what to do about a riot at a football game that threatened to disrupt athletic relations between the two schools. Relatively few students, especially at higher levels of SAT scores, chose these topics, but large numbers of students at all levels of SAT scores chose each of the more "abstract" topics: (1) Who should go to college? (2) When should teen-agers be treated as adults? (Papers on these topics will be referred to as "college" papers and "maturity" papers.) Contrasting points of view on each of these topics were presented as "stimulus material." If these topics are criticized as too "abstract" for young students to treat adequately, the reply must be that they were offered a chance to deal with concrete incidents, close to the lives of students, but they rejected it. The actual wording of the topics that were retained will be found in Appendix C.

Some critics may wonder why we did not have all the papers written on the same topic, since there is some evidence that papers are graded more reliably when all of them deal with the same topic. The answer is that reliability of grading (from one reader to another) was not important in the present research, since it was the differences in grading standards that we wanted to study. We also wanted to find factors that would apply to more than one particular topic. While one cannot generalize from the present study to all topics, a provisional check showed that results for Topic 1 were in reasonably close agreement with results for Topic 2. The consistency of grading of the five readers with highest loadings on each factor was estimated separately for Topic 1 and Topic 2 with the following results:

	Reliability of Ratings				
Factors	I	II	III	IV	V
Topic 1 (papers 1-150)	.45	.30	.28	.41	.36
Topic 2 (papers 151-300)	.49	.20	.27	.48	.33

The method used was the one given by Ebel (1951) for estimating the reliability of ratings. Since these five different groups of readers agreed with one another about as well on Topic 1 as on Topic 2, and since factors are based on such agreements, we concluded that the factors found in this study could be generalized at least to similar types of writing that are commonly assigned in college composition courses.

Out of the 600-odd papers submitted, 300 papers were chosen (without reading the papers) on Topic 1 and Topic 2 solely on the basis of the Scholastic Aptitude Verbal Test scores of the writers so as to include as wide a range in verbal ability as possible without an excessive number of papers at any part of the distribution of scores. The following numbers of papers were chosen at the following levels of SAT Verbal scores:

Number of papers	SAT Verbal scores
63	Below 500
98	500-599
92	600-699
47	700-799

All of the papers written by students with scores below 500 and above 700 were included. Those written by students with scores between 500 and 700 were chosen in a random fashion to secure as even a distribution of scores as possible. The result of this method of choosing papers was that the 300 papers represented a wider spread in verbal ability than any one teacher would be likely to encounter in a selective college, and that the distribution of scores was as nearly "rectangular" as possible (that is, there were about the same number of papers at each level of SAT Verbal scores rather than a concentration of papers with middle scores). It was impossible, of course, to secure as many papers with extreme scores as papers with scores between 500 and 700, but we came as close as the sample permitted. The low correlations between readers' grades could not, therefore, be attributed to the fact that the papers represented a narrow range in verbal ability.

Copies of the 300 papers were typed single-spaced, with all student errors faithfully preserved, and multilithed -- each on a separate sheet to facilitate sorting into piles. Most of the papers were short enough to fit on one side of the page; a smaller number ran over to the back of the page; but none required more than the front and back of one page. Paper number 100 on Topic 1, "Who should go to college?", is reproduced below.

> A college or university is a place where the individual can pursue intellectual interests, it is true. However, it is also a place where the individual develops socially and spiritually as well as intellectually. For this reason, I feel that everyone who truly desires a higher education should have access to those facilities.
>
> Colleges and universities should not be "brain trusts" composed only of those students in the "top third" or "top fifth" of their classes. Today's system of "ranking-" I put the word in quotes because the system, no matter how complex and up-to-date, is not a human one and therefore inaccurate - is highly inefficient. In many secondary schools, the persons who are ranked highest in their class follow business and general education courses. These people have no intention or desire to attend college - and yet they, by ranking so high, displace college-bound students from the "golden circle" as it were. Therefore, I feel that a segregated ranking system should be used in schools that must have a system of this type.
>
> In spite of my personal belief however, I realize that selectivization is the only answer to the present problem of overcrowding in our nation's colleges and universities.

> Unless our colleges and universities prefer to lower their educational standards, which I am sure they do not, then the only answer is the exclusion of those persons who would not benefit from a college education.
>
> Along with this "culling out" of all the deserving students, I believe that all types of athletic scholarships should be discontinued and their funds placed at the disposal of more worthy students - for if college is a place where the mature individuals may pursue knowledge, I hardly think that athletic scholarships are of any importance. The students will be much too busy in the library to attend an athletic event.
>
> In short, selectivization is necessary, but, along with this program, much of the college and universities "drawing power" will necessarily be cancelled out, leaving only the shell - for college is more than just a group of buildings - it is a way of life.

This paper received the following grades from the 53 readers:

Grade	1	2	3	4	5	6	7	8	9 (high)
How many gave it	2	5	12	13	10	8	2	1	0

This response is fairly typical in that more papers (111) received eight of the nine possible grades than any other number of different grades. Since the readers read the papers in their numerical order (an order determined by chance), it may be assumed that this paper was read after the readers had read 99 papers on the same topic. Thus they had an adequate basis for comparison without the boredom and fatigue that might have clouded their judgment of the last paper on this topic, paper number 150.

How the readers were chosen

Readers were nominated by the following persons, who were widely acquainted with outstanding representatives of each field who were known to be deeply concerned with student writing. Their positions at the time the nominations were made are given.

English	Paul B. Diederich, Director of Research in English, ETS.
Social Science	Robert J. Solomon, Head, Social Studies Section, Test Development, ETS.
Natural Science	Frederick L. Ferris, Jr., Head, Science Section, Test Development, ETS.
Law	John A. Winterbottom, Program Director, Law School Admission Tests, ETS.
Writers and Editors	Woodrow Wirsig, Editor, *Printers' Ink*.
Business Executives	Woodrow Wirsig, above, with the assistance of Ralph G. Sharp, Research Director, Street and Smith Publications.

About twice as many representatives of each field were nominated as could be used. Letters of invitation to participate in the study were first sent to representatives of the widest possible diversity in geographical region, institution, position, sphere of activity, and the like. Some who had to refuse were replaced by others of like background. In all, ten representatives of each field agreed to participate, but in a study of this magnitude it was expected that some would be prevented by other duties from completing their assignments. It was gratifying that 53 of the 60 readers completed their assignments, with enough representatives of each field to demonstrate any differences in grading standards that might exist from one field to another. The names and institutional connections of these readers may be found in Appendix D. Those who know these institutions will recognize that these are not ordinary readers; they are the very best group of readers we could assemble who had not been indoctrinated with a "party line" by long service as College Board readers. If their grades do not agree, it is not for lack of interest, knowledge, or sensitivity, but because competent readers with their diversity of background will genuinely differ in tastes and standards.

The numbers given these readers in the various tables do not correspond to the alphabetical order of their names in Appendix D. Hence loadings, correlations, and the like should not be attributed to any particular reader. Regardless of any particular results we feel equal gratitude to all of them. Anyone who has ever submitted to the mortal agony of grading and commenting on 300 papers will give these readers heartfelt thanks.

How the papers were graded

The following excerpts from the instructions to readers present all relevant information on the method of grading:

> "Each reader is to sort the papers into nine piles in order of merit and write on each paper the number of the pile in which he placed it. On as many as possible he should write a brief comment on anything that he liked or disliked about the paper.... We suggest that about fifty of the 150 papers on each topic first be sorted into three piles: average, above average, and below average. Then each of these piles should be sorted on a second reading into three more piles, making nine piles in all. After the first fifty papers have been sorted very carefully into nine piles in order of merit, the hundred other papers on that topic can be sorted rather quickly

> by matching them with the papers already assigned to these piles. If the papers happened to be normally distributed in writing ability, the number of papers finally placed in each pile would be roughly 6, 12, 18, 24, 30, 24, 18, 12, and 6, but there is no reason to expect such a perfectly symmetrical distribution in this small sample. The only absolute rules are that all nine piles must be used; at least six of the 150 papers on each topic must appear in each pile; and the lowest (worst) pile must be numbered 1, the average pile 5, and the highest pile 9.
>
> "Use your own judgment as to what constitutes 'writing ability.' Do not assume that we want you to do this or that. We want you to use whatever hunches, intuitions, or preferences you normally use in deciding that one piece of writing is better than another. You need not even act as a representative of your field, since individuals in any field have varying tastes and standards. Your judgments will be identified only as those of Reader XX, with no name attached in any report. Neither these students nor their teachers will be told how any particular paper was judged."

It should be noted that the readers were not asked to rate the papers on particular qualities, as was done in the two previously reported factor studies of writing by Remondino (1959) and by Carroll (1960), because that procedure appears to yield only clusters of adjectives or rubrics that are logically compatible with one another. (See Appendix I for a criticism of the Carroll study on this point.) In the present study, the factors were based entirely on correlations among grades on "general merit." They were interpreted by categories of comments that were made most frequently by the three highest readers on a factor and far less frequently by the three lowest readers on that factor, as will be explained later.

The distribution of grades given by each reader is given in Appendix F and the distribution of grades received by each student is given in Appendix G. Since these tables are very large, the variation in the grades assigned to a random sample of every thirtieth paper by all 53 readers is illustrated in Table 1.

Table 1

How Many Readers Placed Every Thirtieth Paper in Each Pile

Paper #	1	2	3	4	5	6	7	8	9	Average	Rank
30	2	4	3	8	9	5	11	5	6	5.60	4
60		2	3	2	2	8	10	13	13	6.98	1
90	13	15	13	4	3	4		1		2.74	10
120	1	5	9	10	14	8	4	2		4.53	6
150	3	5	8	16	9	4	8			4.26	7
180	1	1	6	4	14	6	5	8	8	5.92	3
210	4	10	10	14	9	5		1		3.64	8
240			1	7	10	12	9	8	6	6.30	2
270	6	7	21	4	11	3	1			3.38	9
300		2	7	11	17	8	5	2	1	4.94	5

The intercorrelations

The grades assigned to the 300 papers by each reader were correlated with those assigned by every other reader. In this process the three College Board tests, SAT Verbal, SAT Mathematical, and the English Composition Test, were treated as three additional readers who were rating the 300 students who had written papers. The resulting intercorrelations are presented as a massive 56 by 56 table in Appendix B. This table holds in it the information we need on the agreements among the readers and the schools of thought that divide them, but it is a very complicated table, very difficult to interpret without further statistical analysis.

It is possible, however, to see in this big table some important things that have been summarized in Table 2.

1. The median of all correlations between readers is .31. This is often called the "reliability" of the reading. It is very low, indeed, and represents more disagreement than agreement. However, it must be remembered that these readers represent different occupations; any instruction that might predispose these readers in a way that would increase their agreement was carefully avoided.

2. The English teachers had a higher median intercorrelation (.41) with one another than did any of the other occupational groups.

3. All groups agreed with the English teachers better than they did with other members of their own profession.

These three findings suggest that the English teachers, perhaps because of their appropriate training, have succeeded in grading the

Table 2

Median Correlations

Median of all correlations among readers was .31

Median Correlations Among Fields

	English	Soc. Sc.	Nat. Sci.	Wr. - Ed.	Lawyers	Business
English	.41	.33	.35	.30	.39	.30
Soc. Sci.	.33	.28	.33	.25	.31	.25
Nat. Sci.	.35	.33	.34	.26	.34	.28
Wr. - Ed.	.30	.25	.26	.26	.28	.22
Lawyers	.39	.31	.34	.28	.38	.28
Business	.30	.25	.28	.22	.28	.22

Median Correlations Within Fields

English	.41
Lawyers	.38
Nat. Sci.	.34
Soc. Sci.	.28
Wr. - Ed.	.26
Business	.22

Median Correlations With English Teachers

English	.41
Lawyers	.39
Nat. Sci.	.35
Soc. Sci.	.33
Wr. - Ed.	.30
Business	.30

papers more reliably than the other groups. These findings also suggest that any schools of thought with regard to grading that may exist among these readers do not follow occupational lines. This question can be answered by the results of the factor analysis.

The factor analysis

The factor analysis serves to reduce the complexities presented by the table of intercorrelations to the point where interpretation is possible. An informal discussion of factor analysis is given in Appendix J. Some more technical aspects of the present analysis are presented in Appendix A and its accompanying tables. Most pertinent is Table A-2, which lists the "loadings" for each reader on each of six "factors." For purposes of this discussion we can consider a "loading on a factor" to be equivalent to the "correlation with a school of thought."

Factors arise to the extent that certain groups of readers show more agreement with one another than they do with readers who are not members of the group. In the present study more than half of the variability in the grades came from random error or purely idiosyncratic grading standards. Much of the rest of the variability represents agreement on grading the papers. This is brought out most clearly in a later section which discusses correlations with consensus scores. A rather small part of the variability can be described in terms of the five schools of thought that were found by the factor analysis. (The sixth factor, as will be discussed later, is a test factor.) Tables 3 and 4 will give the reader some idea of the kind of differences that produce these factors. These two tables are set up for comparison with

Table 1, using the same ten papers for illustration. Table 3 summarizes grades assigned by seven English teachers with loadings of .30 or above on what will be interpreted as a "Mechanics" factor, and Table 4 summarizes grades assigned by seven readers from five different fields with loadings of .40 or above on what will be called an "Ideas" factor.

Table 3

How Seven English Teachers High on "Mechanics" Graded These Papers

Paper #	1	2	3	4	5	6	7	8	9	Average	Rank
30				1	3		2		1	6.00	3.5
60							3	4		7.57	1
90			5		1	1				3.71	8
120			1	3	2	1				4.43	6
150	1	1		3	2					3.57	9
180					2	2	1	1	1	6.57	2
210		1	2	4						3.43	10
240				2	1	2		1	1	6.00	3.5
270	1		1		4	1				4.29	7
300				1	4	1	1			5.29	5

Table 4

How Seven Readers High on "Ideas" Graded These Papers

Paper #	1	2	3	4	5	6	7	8	9	Average	Rank
30			1	2	2		2			5.00	5.5
60							2	1	4	8.29	1
90	1	2	2		1	1				3.14	9
120	1	1	1	1	2	1				3.71	8
150				2	2		3			5.57	4
180				1	2			1	3	7.00	2
210			1	2	2	2				4.71	7
240				1	2	1	2	1		6.00	3
270	1	1	4		1					2.86	10
300			1	2	2	1		1		5.00	5.5

There were only three papers out of these ten in which there was a discrepancy greater than two ranks in the rank-order assigned by the average of all 53 readers, the seven English teachers high on Mechanics, and the seven readers high on Ideas. Paper 150 was ranked fourth from the top by readers high on Ideas but next to the bottom by readers high on Mechanics; the average of all 53 readers ranked it low--seventh from the top. Papers 210 and 270 were regarded as poor by everyone; the average reader ranked them 8 and 9 out of these ten; but readers high on Ideas preferred 210 to 270, while readers high on Mechanics preferred 270 to 210, and by the same amount: a difference of three ranks.

Such differences may seem too slight and chancy to generate factors; and, indeed, there are only small differences in emphasis between factors On the other hand, three substantial differences in rank order out of ten papers chosen at random are not really "slight." If the differences continued at this rate, there would be 90 such discrepancies in the placement of all 300 papers by the three groups. That would be quite enough to generate the kinds of factors found in the present study.

How the comments were classified

The factors as they emerged from the electronic computer had no names. They showed only that there were certain tendencies toward agreement or "schools of thought" among the readers' grades, but afforded no clues as to qualities in the papers that influenced one "school of thought" more strongly than other schools. Such clues had to be sought in an analysis of the comments that the readers had been asked to write on anything they liked or disliked on each paper.

Since each of the 53 readers had marked 300 papers, there were 15,900 papers to be considered. A tally of comments on 3,557 selected papers (selected without reference to length of comment) indicated that well over 175,000 words had been written on the papers by all 53 readers. Sometimes a single word or an abbreviation, such as "Sp." (Spelling), constituted a classifiable comment.

It took a long time, largely by trial and error, to find out how to reduce this mass of data to manageable form without biasing the interpretation of factors. Perceptible differences began to appear only when we restricted the analysis to the three highest and three lowest readers on each factor, and to all the papers that these readers had placed in their three highest or three lowest piles. There were 3,557 such papers--22% of the total number of papers--which seemed an adequate sample.

If there had been no overlap among the readers, there would have been thirty readers in this sample: three highest and three lowest on each of five factors. As might be expected, however, three of the readers who were highest on one factor were also among the three lowest on some other factor. One reader who was highest on Factor I was among the three lowest on both Factor II and Factor V. After some debate, it was decided to retain these readers in the sample, since they fairly represented both a high position on one factor and a low position on another. Their comments could be classified and counted only once, but the proportions of their comments appear with the high readers on their high factor and with low readers on their low factor (or factors).

This duplication of roles reduced the number of readers whose comments had to be classified from 30 to 25. However, this number seemed to be sufficient for characterizing the sample of 53 readers in certain general ways as well as for representing the factors in this interpretation procedure.

There was also one reader among the three lowest on Factor V who wrote no comments, and another reader among the three lowest on Factor IV who wrote comments only on papers 1-150. Since this phenomenon appeared in the sample of readers with about the same frequency as in the total group of 53 readers, it was decided to retain these readers also. Their position on the factor could not be disregarded without introducing some bias into the sample, and their lack of comments would make little difference in the interpretation of these factors since they did not stand _for_ anything in particular but only for the absence of something. That is to say, the comments of the three lowest readers on each factor were used only as a check on what the three highest readers stood for. If the three highest readers stood first in comments on a certain quality, but the three lowest readers on that factor did also, that quality would have to be discounted in interpreting the factor. It was found that this sort of check could be made by noting the relative position of each quality in the comments of the "low" readers on Factors IV and V who wrote comments.

Out of the sample of 3,557 papers, the following percentages of papers received comments of the following total numbers of words:

Length of Comment	Percentage of Papers
No comment	23%
1 - 8 words	39%
9 - 16 words	20%
17 - 40 words	12%
41 - 100 words	5%
Over 100 words	1%

There was no significant difference in the length of comments between papers graded high (7-8-9) and low (1-2-3). If we code the first three categories of length in the foregoing table as 0, 1, and 2, the average length of comment on the "college" topic was 1.52 for high papers and 1.47 for low. On the "maturity" topic it was 1.30 for high papers and 1.29 for low. If we lump both topics together, the average length of comment was 1.41 for high papers and 1.38 for low.

There was, however, a slight difference between topics: 1.49 for "college" and 1.29 for "maturity." The simplest explanation is that the "maturity" papers were read after the "college" papers, when interest and energy had probably flagged.

A classification of comments was developed by taking readers at random and writing captions under which their comments could be classified until no further types of comments could be found. After many changes, 55 categories of comments were adopted and arranged under seven main headings:

1. Ideas
2. Style
3. Organization
4. Paragraphing
5. Sentence structure
6. Mechanics
7. Verbal facility

Scoring sheets were prepared with the 55 categories of comments listed at the left under these seven main headings, and with several columns in which a reader's comments could be tallied. A sample score-sheet will be found in Appendix H. The score booklet for each reader consisted of twelve score-sheets, one for each of the three lowest scores (1-2-3) and one for each of the three highest scores (7-8-9) for Topic 1, and the same for Topic 2. All of the scoring was done by one person, a skilled proofreader who knew nothing about the standing of the readers on the various factors. Her work was supervised and checked from time to time. Appropriate rules were set down whenever any difficulties arose.

In scoring the papers, the scorer read each comment, recorded its total length, and then entered a tally in all categories covered by the comment. For example, if a reader said of a paper, "Good ideas and organization but several spelling errors," tallies were entered opposite "Quality of Ideas," "Organization," and "Spelling." If the reader went on to encircle two spelling errors, he was given two additional tallies opposite "Spelling." When a comment was unclear or did not fall into any category, it was not tallied. The latter occurred very infrequently.

Analysis of comments by factors

Table 5 shows a tabulation of the comments by readers who are high and low on each factor. The numbers listed under each factor are sums of percentages of comments on Topic 1 and Topic 2 made by the three highest readers on each factor (at the left) and by the three lowest readers on each factor (at the right). Each, therefore, is the sum of six percentages (three readers, two topics) except in Factor IV (where one of the three lowest readers commented on only half the papers) and in Factor V (where one of the three lowest readers wrote no comments). Only the rank-order of the sums was used in interpreting the position of the low readers on these two factors.

Percentages were used rather than raw numbers because some readers wrote so many more comments than others that the raw numbers were not comparable. Each sum of these percentages should be regarded only as an index of emphasis on a given category of comments, since everyone knows that percentages of different-sized wholes cannot be added to yield a percentage of anything. In particular, they do not correspond to the "percentage of all comments" in the last column at the right, which gives the actual percentages of all 11,018 comments that were assigned to each category. These show the relative amount of attention that each

Table 5

Analysis of Comments of Highest and Lowest Readers on Each Factor

Categories of Comments	Three highest readers on						Three lowest readers on						% of all Comments
	I	II	III	IV	V	Total	I	II	III	IV	V	Total	
I. IDEAS	*												
Relevance	44	26	29	15	34	148	16	40	23	20	29	128	3.03
Concise-wordy	20	15	03	17	18	73	22	13	19	24	09	87	2.26
Clarity of ideas	18	10	16	09	15	68	07	35	11	06	16	75	1.84
Quantity of ideas	28	09	01	20	08	66	09	30	13	08	14	74	1.64
Development	19	18	04	12	09	62	13	11	12	03	05	44	1.72
Too brief or long	21	13	02	10	14	60	06	07	01	06	02	22	.58
Persuasiveness	16	00	07	03	06	32	06	20	03	06	06	41	.81
Ending	13	04	03	03	09	32	07	10	06	03	09	35	.76
Generality	09	08	00	03	03	23	00	03	11	00	05	19	.55
													13.19
II. FORM		*											
Spelling	58	105	06	73	72	314	96	11	48	07	31	193	11.82
Clarity of express.	39	65	57	62	25	248	85	58	17	12	33	205	7.06
Organization	24	44	35	09	26	138	15	19	52	23	15	124	4.45
Coherence of ideas	21	34	19	17	29	120	22	32	31	21	17	123	3.33
Reader agreement	10	39	05	23	08	85	24	25	07	16	10	82	1.71
Analysis	13	35	03	04	17	72	01	08	42	19	10	80	1.35
Maturity	07	09	01	08	07	32	06	08	12	05	06	37	1.31
													31.03
III. FLAVOR			*										
Quality of ideas	72	51	89	30	43	285	51	64	51	26	25	217	7.61
Style (general)	19	05	77	08	19	128	32	14	13	31	09	99	4.60
Mechanics (general)	19	01	57	19	00	96	01	26	14	01	07	49	4.37
Originality	15	08	39	13	16	91	07	24	33	05	04	73	2.66
Interest	08	02	32	10	15	67	04	11	06	04	02	27	1.19
Beginning	03	03	20	01	12	39	01	01	05	07	01	15	.61
Sincerity	05	06	12	00	09	32	04	06	04	02	04	20	.57
Info. & Illus.	06	01	08	01	06	22	05	01	10	05	00	21	1.11
													22.72
IV. MECHANICS				*									
Punctuation	10	02	00	69	29	110	17	06	23	15	17	78	6.54
Grammar	17	17	02	39	22	97	34	08	14	17	23	96	4.87
Sentence structure	01	04	00	21	21	47	11	06	03	14	02	36	1.77
Phrasing, idiom	10	03	00	14	05	32	18	04	10	08	18	58	2.52
													15.70
V. WORDING					*								
General	20	35	45	20	49	169	10	52	49	33	04	148	6.04
Word choices	18	07	04	32	34	95	43	13	26	88	44	214	6.43
Logic	10	10	09	05	11	45	03	10	14	01	06	34	1.44
Clichés	06	03	10	04	10	33	06	11	01	02	08	28	.44
Jargon-slang	03	02	00	03	03	11	04	03	00	00	05	12	.38
													14.73
	Omitted minor categories:						2.63% of comments						97.37

---

*The column in which the highest readers on each factor stood first is marked by an asterisk.

category attracted from all twenty-four readers in the sample, but they should not be taken as an index of relative importance. For example, it might be difficult to write more than three or four comments on the ideas in a paper, but it is easy to encircle and comment by abbreviation on many errors in spelling, punctuation, and word-choices. Figures for these categories would be even more inflated if it were not for the tendency of some readers (especially the writers and editors) either to ignore these minor faults or to write only a general comment, such as "poor spelling."

If actual percentages of total comments by the three highest or lowest readers on each factor had been used as the index of emphasis, they would have been unduly influenced by the reader who wrote the most comments. For example, on Factor I, reader 15 wrote 554 classifiable comments; reader 11 wrote 120; reader 25 wrote 269; and the total for all three was 943. Obviously, any percentage based on this total would have been unduly influenced by reader 15, who wrote more than half the comments. We therefore used the percentage for each reader, separately for Topic 1 and Topic 2, and added the six percentages to get our index of emphasis. We computed percentages for Topic 1 and Topic 2 separately to see whether any reader in our sample showed marked differences in emphasis from one topic to the other, but we found no differences greater than might be expected by pure chance.

The purpose of Table 5 was to guide us in the interpretation of the factors that came out in the factor analysis. Therefore, we did not list the categories of comments in any preconceived way that seemed logical to us. Under each factor we listed quite mechanically all categories of comments in which the three highest readers stood first (higher than the readers on any other factor), whether or not the category had any logical connection with other categories listed for that factor. The column in which the three highest readers stood first is marked by an asterisk. In Factor IV, Mechanics, the three highest readers were tied for first place on sentence structure with the three highest readers on Factor V, Wording; hence these two adjoining figures have been underlined. The same was done for the tie on clichés and for the three-way tie for first place on jargon-slang under Factor V. There were no other ties for first place. Categories under each factor were listed in rank order of total emphasis by the fifteen readers who were highest on all five factors.

It may be noted that the smallest "percentage of all comments" for any category included in Table 5 was .38 (meaning .38 of one percent) for "jargon-slang" under Factor V. It may also be noted that this category did not discriminate among the factors, since there was a three-way tie for first place and none of the percentages was significantly different from zero. None of the categories with lower frequencies of occurrence than "jargon-slang" discriminated among the factors and were therefore omitted from Table 5, but they are listed below for the sake of completeness. Each percentage listed is a fraction of one percent of the 11,018 comments on 3,557 high and low papers by the twenty-four readers in the sample.

Omitted Minor Categories	%	
Presents opposing idea	.30	
Amount of illustration	.06	
Originality of expression	.05	
Humor	.14	
Variety	.01	
Middle	.06	
Transitions	.15	
Paragraphing (general)	.55	(noted only by "low" readers)
too brief	.19	
too long	.18	
Variety in sent. struct.	.07	
Type of sentence	.04	
too short	.04	
too long	.18	
Vagueness	.14	
Ornateness	.14	
Word form	.18	
TOTAL	2.63	

Note that the only exception to the cutting-point (below .38) for inclusion in this list was .55 for general comments on paragraphing. These could not be included in Table 5 for the reason that not one of the "high" readers on any factor made any comments of this type; these comments were made only by "low" readers, and hence could not be placed with any factor. In view of the long treatments in the handbooks, this omission (and that of comments on sentence variety) is hard to understand.

Interpretation of Factor I

It was mentioned earlier that all categories of comments in which the three highest readers on each factor stood first were listed under that factor in Table 5 whether or not they had any logical relationship to the name finally adopted for the factor. Thus the reader can judge for himself whether the name fits the data. He should not expect too much. In factor studies of physical or biological data, or of highly reliable educational tests, a factor usually resembles a line with all scores on that factor falling on or near the line. Here, with the extremely low level of agreement among readers -- even among readers who stood highest on the same factor -- the factors were said to resemble "overlapping blobs." It was hard to make any sense of the factors at all, and one should not expect the sort of evidence for each interpretation that would stand up in court. With this general disclaimer, we shall proceed to make what sense we can.

Factors I and IV, Ideas and Mechanics, were the most clearly defined of the five reader-factors in that sixteen readers had loadings above .25 on Factor I and thirteen readers on Factor IV. There were only nine such readers for Factors II and V and seven for Factor III.

The three highest readers on Factor I seemed to be influenced primarily by the ideas expressed: their relevance, clarity, quantity, development, and persuasiveness (or "soundness"), but not to the same extent by other aspects of ideas. "Quality of ideas," for example, was mentioned most frequently by the three writers and editors who were highest on Factor III, Flavor, but the readers high on Ideas were a close second. In general, Factor I called for sound, reasonable, defensible ideas; Factor III for interesting, unusual ideas. One gets the impression that the Factor I readers treated the two topics as examination questions and wanted answers that they could accept as correct and adequate; the Factor III readers wanted salable articles -- articles that would interest readers of popular magazines.

Another discrepancy is that "coherence of ideas" (unity and consistency) was mentioned most frequently by readers high on Factor II, which has been called "Form" -- a purposely vague term intended to refer chiefly to organization and analysis. The three lowest readers on Form, however, put almost the same emphasis (32) on coherence of ideas as the three highest (34).

It may be only a random fluctuation that readers high on Ideas commented most frequently on an effective or ineffective ending, while the writers and editors high on Factor III, Flavor, commented more than any other group on the quality of the beginning. One can see why the latter group would be interested in an attention-getting beginning, but the only reason that comes to mind for the special interest of Factor I readers in the ending is that it frequently sums up the writer's ideas.

It is also easy to see why Factors I and II should be about equally interested in the development of ideas: Factor I, from the standpoint of adequate support for ideas; Factor II, from the standpoint of analysis and organization. Both were about equally averse to generality. In Factor II, this aversion is related to the marked preference of the readers for analysis.

Factor I readers happen to stand first in comments on conciseness vs. wordiness, but this position is not significant, since readers low on this factor made even more comments of this nature and several other groups are almost equally high. They stand out above other groups, however, in comments that whole papers were either too brief (15) or too long (6) for a total index of 21. Their concern with quantity of ideas (fullness of content) and adequate support for ideas probably accounts for this emphasis.

For a time, however, it appeared that Factor I readers might have been influenced chiefly by the sheer length of papers, and that the name for this factor, consequently, should be "Length" rather than "Ideas." The reason for this hypothesis was that a random sample of 30 papers, scored by the five highest and five lowest readers on Factors I and II, was used at one stage as a check on whether certain objectively measurable quantities might be correlated with these two factors. The quantities studied were (1) length, (2) proportion of unusual words (below 50 occurrences in the Lorge magazine list), and (3) number of mechanical errors -- not those on which readers happened to comment but a complete listing of all identifiable errors, done independently by two specialists at ETS. When interesting correlations were found, the measure was repeated on a second random sample of 30 papers to see whether correlations of the same magnitude would reappear. All of them "washed out" on the replication except the correlation of Factor I with length (in inches, since all the papers were typed in a uniform format). On two successive samples of 30 papers each, the correlation of grades of the five highest readers on Factor I with length was close to .8, while the correlation of the five lowest readers on this factor was about -.3.

It was clear, therefore, that Factor I had something to do with length, but after the comments had been classified, counted, and compared, it was decided that these readers were interested in qualities that necessarily involve length but not in length _per se_. They commented more than any other group on "quantity of ideas" (full, thin, feeble, limited, trivial, superficial, sketchy, rich in ideas, discusses aspects only); on "clarity of ideas" (as distinguished from clarity of expression, which was highest for Factor II); on "development of ideas" (good or poor follow-up of ideas, unsupported assertions, inadequate evidence, etc.); on "relevance" (directness, answers or does not answer the question); and on "persuasiveness" (strong or weak arguments). Obviously, to meet these criteria on fairly complex topics a paper needs adequate length. That the readers were adversely affected by "padding" was indicated by the fact that they stood highest in unfavorable comments on "wordiness" and on "too long" (for the paper as a whole), although neither of these figures was significantly higher than for other groups. Thus, it is an error in interpretation to assert that Factor I "is" length because it necessarily involves length.

A surprise in the interpretation of Factor I is that the clearly defined factor formed by the three College Board tests, which are commonly regarded as measures of "general brightness," might have been expected to have its highest correlation with the factor for Ideas, but this was actually its lowest correlation (-.02). (See Table A-3 in Appendix A.) The test factor had practically zero correlations with Ideas, Form, and Flavor, and was positively correlated only with Mechanics (.50) and Wording (.45). While these correlations may be disturbing, it would obviously be too much to expect such tests to predict the quantity, clarity, relevance, development, and soundness of ideas that these students would have on a particular topic, unrelated to school subjects, that was sprung on them without warning -- especially when some devoted much more thought and care to their essay than others. These tests predict how well students can do on the average over the long haul -- not how well they will do on a particular task selected more or less at random.

Factor I had practically zero correlations with the Form factor and with the test-factor. With the other factors its correlations were .27 with Flavor, .25 with Mechanics, and .55 with Wording (see Table A-3). The three highest readers on Factor I had correlations of .49, .52, and .55 with one another. These correlations and similar ones for the other factors are presented in Table 6 on page 34.

Sixteen of the fifty-three readers had loadings of .25 or above on this factor -- more than on any other factor. Although the writers and editors have been mentioned in connection with Factor III, Flavor, it should be noted that just as many (four) writers and editors were high on Ideas as on Flavor, and their average loading was exactly the same: namely, .34. For that matter, two of them had a high loading on Form (.35), but neither stood among the three highest readers on Form.

Interpretation of Factor II

Factor II was called "Form" after some hesitation, since the highest readers appear to be chiefly interested in spelling (105 as the index of emphasis as against 11 for the lowest readers on this factor), and next to that in clarity of expression (as opposed to clarity of ideas, which was highest on Factor I). Emphasis on spelling, however, was high in three factors, Form, Mechanics, and Wording -- probably since it was easiest to mark; it was high, also, among the three highest readers on Ideas (58) but even higher among the three lowest readers on Ideas (96); it was neglected only by the three writers and editors who were highest on Flavor (06), probably because they were accustomed to leaving such minor faults to their proofreaders.

Table 6

Correlations among Three Highest Readers on Each Factor

(Readers are designated as A, B, and C)

I. Ideas

	A	B	C
A		.49	.52
B	.49		.55
C	.52	.55	

II. Form

	A	B	C
A		.33	.36
B	.33		.19
C	.36	.19	

III. Flavor

	A	B	C
A		.35	.21
B	.35		.23
C	.21	.23	

IV. Mechanics

	A	B	C
A		.48	.62
B	.48		.42
C	.62	.42	

V. Wording

	A	B	C
A		.36	.35
B	.36		.37
C	.35	.37	

Clarity of expression, also, did not discriminate well among the factors. It was highest (85) among the three lowest readers on Ideas; it was almost as high (58) among the three lowest readers on Form as among the three highest (65); and these three were followed closely by the three highest readers on Mechanics (62) and on Flavor (57). It was such a vague and easy type of comment (chiefly an underlined expression, followed by "not clear") that it may be discounted in determining the meaning of the factor.

The distinctive emphasis of the three highest readers on Factor II appears to be on organization (44 for the high readers as against 19 for the low), and on analysis (35 for the high readers as against 08 for the low). There is some emphasis on "reader agreement" (expressions of agreement or disagreement with the ideas put forward in a paper), but there is far less difference between high and low readers (39 as against 25, with groups low on Ideas and high on Mechanics close behind with 24 and 23), and it is hard to justify this emphasis. It undoubtedly exists, but students justly complain when their grades are influenced by agreement or disagreement with the ideas of the reader. It has already been mentioned that the three highest readers on this factor mentioned coherence (consistency) of ideas most frequently (34), but this emphasis was negated by the fact that the three lowest readers mentioned it almost as frequently (32). Maturity (of ideas) did not discriminate among the factors, nor did logic, listed under Factor V, which appeared with almost equal frequency among all factors, although one would expect it to be associated with Factor II.

What we are left with, then, to explain the slight tendency toward agreement that marks this factor is an interest in organization and in analysis.

Factor II was one of the three weak reader-factors in that correlations among readers with the highest loadings on this factor were .36, .33, and .19, as shown in Table 6. There were nine readers with loadings above .25 on this factor, of which the largest group was four business executives. This factor had practically zero correlations with all other factors but Mechanics (.33). The latter may be attributable to this group's preoccupation with spelling.

Interpretation of Factor III

Factor III was called "Flavor" because the three highest readers on this factor (all writers or editors) commented most frequently on quality of ideas, style, originality, interest, and sincerity. It has already been mentioned that the readers highest on Ideas, as might be expected, were second in comments on quality of ideas, but no other group approaches this one in proportion of comments on style, which had to do chiefly with the personality expressed in writing (forceful, vigorous, outspoken, personal, inflated, pretentious, etc.) rather than with the word-choices and felicities of expression associated with Factor V. This group happens to be first in comments on originality, which would have supported the interpretation except that the three lowest readers on this factor are a close second (39 to 33). It happens to be first in comments on "Info. & Illus." (information and illustrations) but this position is even more strikingly negated by the fact that the three lowest readers on this factor have more comments of this sort than the three highest. This group is unsurpassed, however, in comments on interest and to a lesser extent on sincerity. It has already been mentioned as an oddity that this group was first in comments on the beginning, while the group highest on Ideas was first in comments on the ending.

It is also odd that this group stood highest in general comments on mechanics, such as "Mechanics poor," but by far the lowest of all groups, either high or low, in specific comments on particular errors. No doubt they leave that sort of thing to the copy-desk editor or the proofreader. They stand second highest in general comments on wording, but this standing is negated by the fact that the three lowest readers on this factor have even more comments of this sort (49 to 45). Again it is characteristic of this group that it has the lowest proportion of comments on specific word-choices. It may be remembered that it was also lowest in comments on spelling.

What appears to generate Factor III, Flavor, is a predominant emphasis on style and interest; a weaker emphasis on sincerity; and an emphasis shared with Factor I readers on the quality of ideas -- ideas that will sell an article rather than ideas that will pass an examination. The last, it must be admitted, is a subjective impression.

Factor III is the weakest of the five reader-factors in that there were only seven readers (four of them writers or editors) with loadings higher than .25; and in that the correlations between grades of the three highest readers were .35, .23, and .21. The factor has correlations of .27 with Ideas, -.03 with Form, .45 with Mechanics, .34 with Wording, and .07 with the tests.

Interpretation of Factor IV

Factor IV was unmistakably a Mechanics factor with predominant emphasis on specific errors in punctuation and grammar. It shared first place in comments on sentence structure with the three highest readers on Factor V, Wording. While it was relatively high in comments on phrasing and idiom, it was surpassed in this category by low readers on Ideas and on Wording. It is significant that seven of the thirteen readers with loadings above .25 on this factor were college English teachers, although only one stood among the three highest readers on this factor.

Factor IV was a strong factor in that it was second highest (after Ideas) in the number of readers with loadings above .25, and in that the correlations among the grades of the three highest readers were .62, .48, and .42. It was closest to Factor V, Wording, in that the correlation between them was .63 (possibly because many of the word-choices that attracted comments were errors in usage). It also had the highest correlation with the test-factor (.50), while almost anyone would suppose that the test-factor would have its highest correlation with Ideas, which turned out to be its lowest (-.02). Since everyone is sensitive in some degree to mechanical errors, or since mechanical errors are relatively obvious and easy to detect, this factor correlates with everything: Ideas, .55; Form, .33; Flavor, .45; Wording, .63; and the tests, .50.

Interpretation of Factor V

It is hard to know what to make of Factor V, which we have called "Wording." It is undeniably a factor, since nine readers turned up with loadings of .25 or higher on it, but it was close to the chance level, and its interpretation is uncertain. Three of the categories of comments in which it is highest or tied for first place did not discriminate among factors but are spread almost equally among all. That leaves only general comments on wording ("felicity of expression") and comments on specific word-choices. In the first it was surpassed by the three lowest readers on Form and equalled by the three lowest readers on Flavor. In the second it is far outclassed by the three lowest readers on Mechanics, by readers who are low on the same factor, Wording, and by readers who are low on Ideas. The only character of its own is given by the fact that there is a greater disparity in general comments on felicity of expression (49 for high readers and 04 for low) than in any other category. Perhaps it is a factor for general verbal facility. It is sharply distinguished from the general comments on style in Factor III, for they had to do with the personality expressed in writing -- forceful, vigorous, outspoken, personal, etc. -- while the comments on Factor V had to do with the wording.

Factor V correlates .63 with Mechanics, .55 with Ideas, .45 with the tests, and .34 with Flavor. Its correlation with Form is negligible. The grades of the three highest readers on this factor were correlated .37, .36, and .35 with one another.

Those who have long memories of objective tests in English composition will remember that in several batteries there were two tests: one dealing with the detection and sometimes the correction of errors, the other emphasizing the choice of words or arrangements of words to fill in blanks in sentences. These tests were called, respectively, Correctness and Effectiveness of Expression, or some variant of these names. "Effectiveness" sometimes included scrambled sentences, outlining, and anything else that would distinguish it from the test on Mechanics, but these miscellaneous exercises usually had low correlations with total test scores, with external criteria, and with other objective tests of the same types. What was central and valid in the "Effectiveness" tests was usually the choice of words or word-orders to fill in gaps in sentences.

It is tempting to think of these two types of objective tests as corresponding roughly with our Factor IV, Mechanics, and Factor V, Wording. The tests usually correlated about .6 with one another, and the correlation of these two factors with one another in the present study was .63. The three objective tests used in the present study also had their highest correlations with these two factors: .50 with Mechanics and .45 with Wording, while their correlations with Ideas, Form, and Flavor were negligible.

We might have called Factor V "Effectiveness" but for the fact that the old "Effectiveness" tests were so frequently contaminated with short exercises on scrambled sentences, outlining, and the like, which were not emphasized in the comments of readers high on Factor V. We finally chose the name "Wording" because the distinctive emphasis in these comments was on the choice and arrangement of words.

Interpretation of Factor VI, the tests

The Scholastic Aptitude Verbal and Mathematical Tests and the English Composition Test of the College Entrance Examination Board are too well known to require explication here. These three tests emerged from the analysis as a single "test-factor," quite distinct from any of the "reader-factors." The three tests had loadings of .67, .65, and .58 on the test-factor and no loading higher than .04 on any of the reader-factors. The test-factor had practically zero correlations with Ideas, Form, and Flavor, and was positively correlated only with Mechanics (.50) and with Wording (.45).

Summary of interpretations of factors

Since the interpretation of the five reader-factors may be blurred by the volume of conflicting evidence that has been presented, it may be advisable at this point to "tick off" the dominant emphases of these factors as we see them now:

I. Ideas: relevance, clarity, quantity, development, persuasiveness.
II. Form: organization and analysis.
III. Flavor: style, interest, sincerity.
IV. Mechanics: specific errors in punctuation, grammar, etc.
V. Wording: choice and arrangement of words.

It is likely that Factors IV and V can be measured by objective tests well enough for a practical judgment, but we see no way at present to measure Factors I, II, and III reliably, either by objective tests or by essays.

Discussions of objective tests sometimes use the premise that "they cannot measure creativity" to imply that essays can. It is worth noting that the adjective "creative" or any close synonym for it did not enter into the comments of these 24 readers frequently enough to count in the interpretation of any factor. The nearest approach to it was "originality," mentioned most frequently by both high and low readers on Factor III, but any comment that an idea was unusual--no matter how bizarre--would be tallied under this heading. It was not emphasized by readers high on Ideas.

It is interesting to observe that the factor analysis has confirmed the tentative conclusion we made after looking at the intercorrelations, the conclusion that the factors do not run along occupational lines: we find no English factor, lawyer factor, etc. The readers with the highest loadings on each factor in all cases represent different fields with one exception: the three readers with highest loadings on the factor called "Flavor" (style, interest, sincerity, etc.) were all writers or editors; no other reader had a loading on this factor higher than .27. Another near exception was that seven of the ten English teachers had loadings of .30 or above on the Mechanics factor, while only three other readers had loadings this high; but the three highest loadings on this factor represented English, Business, and Natural Science--in that order. Table 7 below summarizes the fields for the readers having high loadings on each factor.

Table 7

The Fields of Readers Having Factor Loadings of .25 and Above

(Three highest loadings underlined)

FACTOR I, Ideas

English	.33, .33, .30, .27
Soc. Sci.	.58, .53, .25
Nat. Sci.	.47, .34
Wr. - Ed.*	.40, .37, .31, .27
Law	.41
Business	.44, .43

FACTOR II, Form

Nat. Sci.	.42, .31
Wr. - Ed.	.35, .35
Law	.37
Business	.37, .32, .29, .29

FACTOR III, Flavor

Soc. Sci.	.26
Nat. Sci.	.27
Wr. - Ed.	.41, .34, .33, .28
Law	.26

FACTOR IV, Mechanics

English	.47, .36, .35, .32, .31, .30, .
Soc. Sci.	.26
Nat. Sci.	.38
Wr. - Ed.	.29
Law	.33, .26
Business	.40

FACTOR V, Wording

Soc. Sci.	.35, .29
Nat. Sci.	.29, .25
Wr. - Ed.	.26
Law	.30, .25
Business	.31, .27

FACTOR VI, Tests

SAT - V	.67
ECT	.65
SAT - M	.58

*Wr. - Ed. = Writers and Editors

Correlations with "consensus scores"

A "consensus score" was found for each paper by adding the grades given it by all 53 readers, arranging these cumulative totals in rank order, and converting to "stanine scores" (on a scale of 9 points, with a normalized proportion of papers at each "stanine"). It was then possible to compute the correlation of each reader and each test with these "consensus scores" with the results shown in Table 8.

Table 8

Correlations with "Consensus Scores"

		Fields						
Correlations	Readers	Eng.	Soc.	Sci.	Law	Editors*	Writers*	Business
.66 - .73	10	3	1	2	2	2	-	-
.58 - .63	14	4	3	2	3	-	-	2
.52 - .57	14	3	2	3	3	2	-	1
.36 - .49	12	-	2	1	-	1	5	3
.22 - .31	3	-	1	-	1	-	-	1

.34 Scholastic Aptitude Verbal Test

.33 College Board English Composition Test (semi-objective)

.19 Scholastic Aptitude Mathematical Test

---

*Editors and Writers are listed separately in Table 8 because their correlations with the consensus differ.

The application of a correction for the fact that each reader's grades constituted 1/53 of the "consensus" would lower all of these correlations by a few points. The uncorrected correlations of all readers with this "consensus" will be found in Appendix E.

These results suggest that, if it is necessary to use essay grades on general writing ability with sufficiently high reliability for a practical judgment, the first step must be to test the readers. A by-product of the present study is a collection of 18 papers for this purpose: one paper on each topic from the middle of each "stanine" unless the middle paper was unusually short or long; in which case the nearest paper of average length was selected. The directions to the prospective readers would be to rank the papers on each topic in order of general merit. If one selected only readers whose average rank-order correlation with the consensus of these 53 readers was .60 or better, the inter-reader reliability of the sum or average of four of their grades on the same set of papers of this sort would probably be in the neighborhood of .80.

This, however, would not be the over-all reliability of the grade on writing ability, which must take account of the variability of the student from one topic or occasion to another. In a research study, it would probably take six papers on different topics, each written in a separate testing session, and each graded independently by four tested readers, before the over-all reliability of the grade on writing would reach .90.

Since this would cost about $100 per student at present rates, it would not be practical for the purposes for which essay grades on writing ability are generally used. The nearest approach that would be feasible would be to have each student write a paper in class on two successive days at four different times during a course in writing, and to have each paper graded independently by two tested readers. The over-all reliability of the sum or average of these eight grades would probably reach or exceed .80--low by the standards of objective tests but good enough for most practical purposes.

Effort to find "differentiating papers'

An effort was also made to find sets of papers that would pick out readers who were high on the various factors, but this effort failed. The eighteen papers that had been selected out of the middle of each stanine proved to be unsuitable for this purpose, since the average grade of the three highest and three lowest readers on each factor ranked these papers in pretty much the same order as the consensus, with only such exceptions as might have occurred by chance.

Next, an effort was made to find "high differentiating papers" and "low differentiating papers" defined as follows:

High

(1) At least one of the three highest readers on a factor grades it 9; neither of the others lower than 6.

(2) No high reader on any other factor grades it higher than 7.

Low

(1) At least one of the three highest readers on a factor grades it 1; neither of the others higher than 4.

(2) No high reader on any other factor grades it lower than 3.

It turned out that there were only three papers that met these standards out of all 150 papers written on Topic 1, and so the effort was not carried beyond that point. This whole exploration showed that it was slight average differences in grading all 300 papers rather than sharper differences on any smaller set that generated the factors found in this study.

This exploration could be carried further by taking the factors two at a time and attempting to find papers that were graded quite differently by the highest readers on each factor. These preliminary results, however, did not warrant the expectation that a sufficient number of such papers would be found to make a reliable test. They also made it seem likely that the slight natural bias of a reader toward one factor or another when he was told only to "use his own judgment" might be overcome by directing and training the "generally good" readers to score the papers for such qualities as ideas, organization, and style. That is, there seemed little point in trying to find readers with a natural bent toward one or another of these qualities in an unstructured situation when a few directions and a short period of training could focus attention on these qualities quite specifically.

There remains the possibility, of course, that a reader who is naturally high on ideas will not be able to score papers for organization or for style, even when he is directed to do so. That is, his grades will still be primarily influenced by the ideas, even when he thinks he is grading for organization or for style. The slight differences between factors found in this study, however, does not make this possibility seem very likely. In general, all readers commented on everything; and it was only by resolute analysis that slight differences in emphasis were found. We had carefully refrained from even mentioning such terms as ideas, organization, or style for fear of channelling comments in these directions. If we wanted to train readers to look for such qualities, there is no evidence in the present study that it cannot be done; but it would take another study to show how much differentiation in grades such directions and such training could produce.

Low correlations of test scores with consensus

Those who remember previous College Board studies in which the Scholastic Aptitude Verbal Test and the English Composition Test have regularly predicted average grades in composition courses better than a single essay may be surprised by the relatively low correlations of these tests with "consensus scores" in the present study: .34 for SAT Verbal and .33 for the English Composition Test, while only three of the 53 readers had lower correlations with the consensus than these.

On sober reflection, however, there is no real inconsistency between these results and those of previous studies. These tests will predict how well students will do on the average, over the long haul, better than any single essay -- even when it is graded by expert College Board readers who have been trained to agree on certain qualities. It was never claimed that they would predict how well students would write any one particular paper. Readers who have seen a particular set of papers can naturally judge which students wrote well or poorly on the topic assigned somewhat better than tests of quite a different sort that were taken eight months earlier. If these students wrote twenty more papers, one would then expect their average grades to be somewhat closer to their test scores than to their grades on this single essay.

Correlations with objectively scorable quantities

Flesch (1948) has developed measures of what he calls "reading ease" and "human interest." The reading-ease measure, based on syllable/word and word/sentence ratios, was designed to predict "the average grade level of a child who could answer correctly three-quarters of the test questions asked about a given passage" (1948, p. 222). The human interest measure, based upon the percentage of "personal words" and "personal sentences," was intended to measure "the reader's attention and his motivation for continued reading."

The correlations of reading ease, human interest, and their constituent measures with "consensus scores" on a random sample of thirty of the papers used in this study are shown below.

Syllable/word	.21
Word/sentence	-.06
Personal words	-.17
Personal sentences	.12
Reading ease	-.10
Human interest	-.17

None of these correlations is significantly different from zero. We may conclude that none of these variables is an important determinant of the readers' grades.

The syllable/word and word/sentence measures were found to be correlated -.23 in this set of essays, while Flesch reports a correlation of .46 between these measures in his materials. Perhaps this difference reflects the difference between edited professional and non-edited non-professional writing.

It was previously reported in connection with the interpretation of Factor I that the only other objectively scorable quantity associated with any factor in repeated samples was length; and evidence was cited that readers high on Factor I value qualities that require length but react adversely to unnecessary length.

Leniency-severity

In a final attempt to find objectively scorable quantities that might be associated with the various factors, the percentage of the 300 papers given low grades (1-2-3) and high grades (7-8-9) by the three highest readers on each factor was computed with the following results:

Factor	Low Grades	High Grades
I	28%	22%
II	27%	22%
III	35%	22%
IV	30%	21%
V	21%	18%

It may be remembered from the directions to readers that they were free to assign as many or as few papers as they liked to the various piles, with only the restriction that all nine piles must be used and not less than 4% of the papers must appear in any one pile.

The percentage of papers given high grades by the three highest readers on each factor obviously does not vary more than might be expected by chance. There is a slight tendency for readers high on Factor III to use more low grades than the other groups and for readers high on Factor V to use more middle grades, but these differences are not significant because there was a far greater variation within factors than between factors. In the most extreme case, the three highest readers on Factor III gave 5% more low grades than the next most severe group (Factor IV), but one of these readers gave 116 low grades, another 60, and the third 137 (out of 300 papers) -- obviously a much greater variation. We may therefore conclude that a tendency toward leniency or severity in grading does not distinguish between groups of readers who were high on the various factors.

Discussion and conclusions

Five different types of readers were found by mathematical analysis of tendencies toward agreement in their grades on 300 papers written by liberal arts freshmen of widely varying ability. These five types were identified by the proportions of their comments on "anything they liked or disliked in each paper," as classified and counted by a skilled proofreader who did not know the standing of any reader on any factor. The most clearly defined and sharply contrasting types were 1 and 4 , heretofore designated as "Ideas" and "Mechanics." Readers of Type 1 stand first in comments on certain aspects of ideas: their relevance, clarity, quantity, development, and persuasiveness (or "soundness"). In general, they appear to have a question-answering concept of an essay topic; they want answers that they can accept as relevant to the question, correct, and adequate. As a necessary consequence, grades of the highest readers on this factor were highly correlated with length of paper, while those of the lowest readers on this factor were negatively correlated with length. Their high proportion of unfavorable comments on "wordiness" and "too long" indicates, however, that they would react adversely to "padding."

Their highest proportion of comments was on relevance, which also showed the greatest disparity between readers high and low on this factor. The most typical comment tallied under this heading was "He did not answer the question." This emphasis points to a possible weakness in "grading primarily for ideas" in an essay designed to test writing ability. It is not uncommon in such examinations to find a few papers that readers high on other factors regard as "quite well written," but that readers of this type flunk. Their most common explanation is, "He did not answer the question." The fact that he answered some other question --

possibly a closely related question -- quite well does not weigh with these readers; they want an answer to the question set by the topic as they conceive it. They are not attracted by unconventional ideas; such ideas received more favorable comments from readers of Type 3, the writers and editors. While the evidence found in this study does not indicate that they are repelled by unconventional ideas, their favorable comments indicate that they prefer ideas to be correct, sound, and adequately supported. A rather dull but workmanlike paper of adequate length could earn a high grade from these readers but not from the readers of Type 3.

Readers of Type 4, Mechanics, are equally unambiguous in their preferences; their grades are inversely related to the number of their comments on mechanical errors -- with some adjustment for excellence in other respects. They have by far the highest proportion of comments on specific errors, even though readers of Type 2 surpass them in comments on errors in spelling. Seven of the ten English teachers belong to Type 4.

The other three types of readers are harder to characterize, for there were fewer readers with high loadings on these factors; their correlations with one another were lower; and their comments were more difficult to interpret. Readers of Type 2 place their most distinctive emphasis on analysis and organization. They appear to be obsessed by spelling, but this emphasis may be an accident of the present sample of readers. They also stand high in comments on clarity of expression and coherence or consistency of ideas, but these emphases are less distinctive since they are closely followed by other groups among both high and low readers. One would expect them to stand high on logic, but specific comments on logic appeared with almost equal frequency in several factors.

Readers of Type 3 would be recognized in any group grading an examination as the devotees of creative writing. Their most distinctive emphases are on style, interest, and sincerity. They share with readers of Type 1 an interest in the quality of ideas but tend to prefer the unusual, unconventional ideas to the merely correct. Their high standing on originality is negated by the fact that the three lowest readers on this factor almost tie them. They have the highest proportion of general comments on mechanics, but make up in this way for the fact that they have by far the lowest proportion of comments on specific errors in punctuation, grammar, spelling, or word-choices. It may be significant that four of the seven readers with loadings of .25 or above on this factor were either writers or editors, and no other reader had a loading on this factor higher than .27.

The only distinctive emphasis that could be found in the comments of readers of Type 5 was on the choice and arrangement of words. They are close to the readers of Type 4 in that there is a correlation of .63 between these two factors -- the highest correlation between factors found in this study. Since this is also the usual correlation between objective tests of Correctness and Effectiveness of Expression, it was suggested that this factor might correspond to Effectiveness -- with its usual emphasis on the choice and arrangement of words to fill in blanks in sentences. Sometimes this emphasis is called "style," but it differed sharply from the general comments on style of the Type 3 readers, which were comments on the personality expressed in writing.

It was not unusual or unexpected to find these five types of readers; they would be recognized in any large group of readers grading an examination. What was unexpected was that there would be just these five groups and no others, and that they would hang together in the particular ways that were here discovered.

By contrast, when the College Entrance Examination Board spent six years and $97,700 in the development and tryout of its General Composition Test (1951 through 1956), the distinguished readers who graded it agreed to grade the papers on five qualities: Mechanics, Style, Organization, Reasoning, and Content. Such an analysis cannot be dismissed as "subjective"; it was based on a great deal of knowledge of qualities emphasized in textbooks and of the teaching tradition.

Our mathematical analysis of the unstructured behavior of 53 readers comes up with three of the same factors that we identified as Ideas (rather than Content), Form (rather than Organization), and Mechanics. Thus far we are better off only in knowing more precisely what the readers high on Ideas looked for, and what pitfalls they fell into, while the term "Content" is relatively vague. It might be taken, for example, to refer to the quantity of information presented, but we know that our "Info. & Illus." category was given a very weak emphasis by readers who were high on a different factor. We also know that the heading "Organization" must be broadened to include "Analysis," for the readers who emphasized the first were equally interested in the second. Mechanics remains Mechanics in both studies.

Now we come to two rubrics in the General Composition Test that must be redefined. "Style" breaks up into two parts: comments on the personality expressed in writing (forceful, vigorous, outspoken, personal, inflated, pretentious, etc.) that were emphasized by one group of readers, and comments on the choice and arrangement of words -- on general felicity of expression -- that were emphasized by another. The General Composition Test appears to have considered only the second of these two factors and to have neglected the first, as well as all the other distinctive qualities that were subsumed under our heading, "Flavor." This was probably deliberate, since these types of reader-reactions are associated with "creative writing," which was under a cloud at that time. But here they reappear, not in the mouths of schoolmarm "creative writers" but of successful writers and editors. It may be time to reconsider the validity of their position.

The heading "Reasoning" that was used in grading the General Composition Test corresponds most precisely to our category "Logic," listed under Factor V, which accounted for only 1.44% of the 11,018 comments classified and was spread almost equally over all factors but Mechanics. In a larger sense, it breaks up into three parts. If you are looking for relevance, clarity, quantity, development, and soundness of ideas, you will find them emphasized by readers of Type 1. If you are looking for analysis and consistency of ideas, you will find them emphasized by readers of Type 2. If you are looking for interesting, unusual ideas, you will find them emphasized by readers of Type 3. Thus there are three quite different aspects of "Reasoning," each emphasized by a different type of reader.

The foregoing analysis of a grading system that was developed by the College Board and then abandoned illustrates the main type of "conclusion" that the present study can support. It was a theoretical study; it had no practical application immediately in view; and its data can be used chiefly in thinking about the soundness of grading systems in the field of English composition

To be sure, it analyzed only what 53 competent and devoted readers of varied backgroundsdid in an unstructured situation. It cannot pretend to know what they should have done; hence, in a sense, it has no normative significance. Still, one can reason from their behavior to the type of conclusion just drawn: that "Reasoning," for example, is not a psychologically meaningful category for grading English compositions. It is either one small, relatively unnoticed facet of skill in writing, or it is three quite different things, each emphasized by a different group of readers. "Style," similarly, is not one thing but at least two things: one having to do with the choice and arrangement of words, the other with the expression of personality in writing.

Two other conclusions are suggested by the data:

1. Any reliance upon essay grades on writing ability without computing their reliability is almost certain to be unwarranted. Witness the fact that 94% of these papers received either seven, eight, or nine of the nine possible grades and no paper received less than five different grades from the 53 readers. The median correlation between the readers' grades was .31.

2. The task of readers may possibly be simplified by restricting their attention, insofar as possible, to Ideas, Form, and Flavor (and their constituents). We found nothing in the comments of these readers on Mechanics and Wording that could not be measured more reliably and systematically by objective tests. These two factors take up most of the time in grading, since they offer the largest number of small, specific errors to correct. If they were taken over by objective tests, and if studies were made that would lead to more precise observations of the remaining factors, some progress might be made toward agreement on common standards.

References

Carroll, J. Vectors of prose style. In Sebeok, T. A. (Ed.), *Style in language*. New York: Technology Press and Wiley, 1960, 283-292.

Ebel, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, *16*, 407-424.

Flesch, R. A new readability yardstick. *J. appl. Psychol.*, 1948, *32*, 221-233.

Huddleston, Edith M. Measurement of Writing Ability at the College Entrance Level. *J. exp. Ed.*, 1954, *22*, 165-213.

Olsen, Marjorie. Summary of main findings on the validity of the 1955 college board general composition test. Statistical Report 56-9. Princeton, N. J.: Educational Testing Service, 1956.

Remondino, C. A factorial analysis of the evaluation of scholastic compositions in the mother tongue. *Brit. J. educ. Psychol.*, 1959, *29*, 242-251.

Swineford, Frances. College entrance examination board general composition test, form L. Statistical Report 56-45. Princeton, N. J.: Educational Testing Service, 1956.

Torgerson, W. S., & Green, B. F., Jr. A factor analysis of English essay readers. Research Bulletin 50-30, Princeton, N. J.: Educational Testing Service, 1950.

Acknowledgments

Thanks are due to Mrs. Donna Gunderson for classifying the readers' comments and to Dr. Paul I. Jacobs for estimating the reliability of grading on the two topics (page 5) and for computing correlations with objectively scorable quantities (pages 30, 31, 32, 49, and 50). Dr. Ledyard R Tucker, formerly of ETS, now at the University of Illinois, suggested the method of factor analysis that was used in this study. The Carnegie Corporation of New York provided a substantial research grant to support the study.

Appendix A

Further Details of the Factor Analysis

The most usual kind of factor analysis, often referred to as "R-technique," is an analysis of the intercorrelations among scores on variables, where the variables have been given to a population of people. Another kind of factor analysis, called "Q-technique" or inverse factor analysis, is an analysis of the intercorrelations among the people, where the people have taken a "population" of tests. In the present study we have intercorrelations among people. If we can consider each of the 300 essays to be a "test" of the reader's grading techniques, the present study may properly be called an inverse factor analysis.

The table of intercorrelations among people and test scores appearing as Appendix B was factored to a principal components solution *via* Eigenroots and vectors. The IBM machine computation was carried out at the Littauer Statistical Laboratory, Harvard University.

The roots were plotted in order to make a decision on the number of factors that would be worthwhile using for rotation and interpretation. The first 12 of the 56 roots are as follows:

1.	17.19	5.	.84	9.	.57
2.	3.11	6.	.79	10.	.48
3.	1.38	7.	.68	11.	.47
4.	1.20	8.	.60	12.	.43

From these values the roots declined evenly to -.31 for root No. 56. The picture here is one of very heavy emphasis on the first root, suggesting that beyond the large amount of error in the reading, there is a large central core of agreement among the readers. The rapid decline in the size of the roots toward what may be interpreted as a chance level indicates that the schools of thought operating in the reading are small in number and not striking in importance as compared with the amount of general agreement.

Rotations of factors were carried out graphically to attain an oblique simple structure without significant negative loadings. Because of the pattern of roots listed above, rotations were started with only the first four factors, but it became apparent that a satisfactory structure would be difficult to obtain. Factors 5 and 6, for which the roots seemed to be at least a little above the chance range, were then added and a more satisfying rotation became possible. Note that the addition of Factors 5 and 6 contributes somewhat to the loadings for all of the final six rotated factors; there is no indication here that any two particular factors would have been cut off. The effect on the results of having four rather than six factors would appear as a merging of some of the factors we found and some rearranging of the loadings. Table A-1 gives the loadings for the 53 readers and 3 tests on the first six orthogonal factors before rotation. Table A-2 gives the loadings after oblique rotation.

Although rotation was carried out to oblique simple structure, and the resulting factors are the ones that have been discussed in this report, some research workers, particularly those in England, would probably have allowed the first factor extracted to stand as a general factor, and would have rotated the other five to find what meaningful schools of thought remained in them. The resulting interpretation would have put deserved emphasis on the general factor of agreement among the readers, but the other differences would have been less clear.

Torgerson and Green (1950) carried out a similar study of English essay readers, but their study involved only 38 essays read by 20 readers who had been instructed in how to grade essays on the particular topic being graded. Their method of rotation was to set the large general factor orthogonal to the three group factors that they found. As might have been expected, however, they were not able to interpret the group factors, although one of them seemed to be related to the sex of the readers. We found no factor related to sex.

It is of some interest to look into the rotations of factors in the present study more deeply by inspecting the plots of the factor loadings. Figure A-1 shows some of these plots. (In each plot, the Scholastic Aptitude Verbal and Mathematical Tests and the English Composition Test are indicated by V, M, and E, respectively.) The first row of three plots shows the pairings of the first three factors before rotation: 1 with 2, 1 with 3, and 2 with 3. These three plots taken together depict how the readers and tests lie in the 3-dimensional space described by the first three principal component factors. All readers and tests have a substantial loading on Factor 1, the general factor. The plot of Factor 1 with Factor 2 shows the College Board tests lying out in one sector, with readers 9, 40, and 43 coming closest to them. Readers 11, 15, and 34 apparently find something quite different in the essays. The plot of Factor 1 with Factor 3 shows a difference between readers 27, 40, and 48 on the one hand and readers 11, 17, and 34 on the other. Rotation, even such as to retain the general factor, would simplify this picture somewhat, but the factors would remain bi-polar (having both positive and negative loadings of significant size) and would be hard to interpret.

The second row of plots depicts these factors rotated for oblique simple structure as described earlier and discussed in the body of this report. They are now called Factors I, VI, and II, the Ideas, Tests, and Form factors, respectively. The clear separation between the tests and the readers is very apparent. The lack of significant negative loadings is a welcome simplification, but the plot of the Ideas and Form factors shows how a reader can adhere to either group or can lie anywhere between the two. It is, of course, not unreasonable that a reader such as No. 37 can be found, for whom ideas and form are equally important in evaluating an English composition.

Consider now the columns in Table A-1 headed "h^2." These are the "communalities" for each reader and test. They are actually the sum of the squares of the loadings for all six factors. A communality represents the proportion of the variability in a reader's scores that is explained by his loadings on the six factors. For a large majority of the readers, less than half of the variability in the scores is explained by the factors. The remainder of the variability is brought about by unique ideas on scoring that the reader does not share with anybody or by just plain error. Even if we had found only one general factor of agreement among the readers, this amount of uniqueness and error would explain most of the distressing diversity of grades that were assigned to each paper and the low average correlation between readers. The portion of this remaining variability that is purely error would reveal itself, only too clearly, if we could let the readers grade the same essays a second time in a year or so, after they had completely forgotten the grading standards that they used on the present occasion. We would probably find that the readers would not agree very well even with themselves.

Table A-3 gives the intercorrelations among the factors, which have been discussed to some extent in the body of the report in connection with the interpretation of each factor. It is by inspecting this table that we can learn something about the "general factor" of reader agreement which was discounted by the particular kind of rotation of factors that was chosen. Inspection of the table reveals that we have a table of rank one. That is, if we were to factor analyze this table of intercorrelations, we would find only one "second-order" factor. The over-all agreement of the readers and tests that is not explained by the six factors already discussed appears to be concentrated on one rather than several major aspects of writing. It might have turned out, for example, that Ideas, Form, and Flavor were highly correlated with each other but not with the highly correlated mechanical factors. This would have produced an intercorrelation table of rank two, but this did not happen. Instead, there is a central core formed by Mechanics and Wording, with Tests coming close to it. Ideas and Flavor are farther out from the center, and Form is still farther off by itself.

Table A-1

Unrotated Orthogonal Factor Loadings

Reader	Factor 1	2*	3	4	5	6	h²
English Teachers							
1	.72	-.06	-.12	-.06	.14	.14	.58
2	.57	.02	-.09	-.20	-.12	.09	.40
3	.64	.09	.02	.10	.28	-.01	.50
4	.69	.20	.12	-.09	.10	-.07	.55
5	.54	-.36	-.16	.04	.00	-.07	.46
6	.64	-.20	.07	-.17	.06	.22	.53
7	.71	.09	.15	-.10	.00	.07	.55
8	.59	.14	.09	-.09	.12	.11	.41
9	.63	.23	.15	-.16	.09	.23	.56
10	.56	-.08	.08	.14	.04	.04	.35
Social Scientists							
11	.56	-.37	.24	.16	-.07	.13	.56
12	.26	.01	.01	-.10	-.07	.06	.08
13	.43	-.22	-.11	-.08	-.17	-.08	.28
14	.59	.11	-.08	.11	.04	-.15	.40
15	.47	-.49	.04	.23	-.05	.08	.53
16	.61	.04	.04	-.13	.14	-.09	.42
17	.71	.03	.27	-.06	-.01	-.10	.59
18	.52	-.09	.02	.21	.11	-.10	.35
19	.54	.17	-.07	.03	.02	-.21	.37
Natural Scientists							
20	.73	.12	.16	-.03	.11	.11	.60
21	.52	.04	.09	.02	-.13	-.23	.36
22	.60	-.18	-.14	.07	.16	-.12	.45
23	.64	.13	-.14	.06	.11	-.11	.48
24	.54	-.02	-.22	-.03	-.14	-.09	.37
25	.72	-.27	-.03	.19	-.11	.04	.65
26	.48	-.35	.17	.01	-.12	.02	.40
27	.54	.02	-.43	-.01	.12	.00	.50
Writers and Editors							
28	.53	-.14	.00	-.04	-.09	.00	.31
29	.32	.00	-.16	-.35	-.23	.10	.31
30	.66	.01	.16	-.15	-.11	-.02	.50
31	.48	.18	.07	.08	-.09	-.13	.30
32	.69	-.08	.04	.01	.16	.04	.51
33	.35	-.40	-.04	.13	-.04	.01	.31
34	.46	-.39	.24	-.13	-.21	.06	.49
35	.57	.11	.17	-.13	-.20	-.19	.46
36	.46	-.01	.14	.03	.03	-.15	.25
37	.39	-.20	-.28	.05	-.06	.25	.34
Lawyers							
38	.61	.15	-.04	-.15	.04	.05	.42
39	.64	.19	.02	-.13	-.05	.08	.47
40	.20	.26	-.34	-.14	.00	.11	.26
41	.73	-.21	.17	.10	.02	.07	.61
42	.53	.11	.03	.00	-.21	-.15	.37
43	.57	.37	.06	-.22	.05	.01	.52
44	.69	-.05	-.14	.00	-.09	-.08	.52
45	.61	.20	.12	-.04	.11	-.18	.47
46	.59	.04	-.01	.01	.14	-.19	.40
Business Executives							
47	.42	-.06	-.27	-.13	-.06	.14	.29
48	.40	.21	-.33	-.01	.03	-.08	.32
49	.55	-.19	-.16	.26	.18	.07	.47
50	.63	-.01	-.22	.06	-.05	.01	.45
51	.48	.15	.17	-.09	.13	.16	.34
52	.28	.02	-.20	-.15	-.19	-.12	.19
53	.57	-.32	-.08	.18	-.10	.00	.48
Tests							
SAT-V	.39	.62	.04	.39	-.17	.12	.74
SAT-M	.23	.53	.01	.29	-.21	.15	.48
ECT	.39	.62	-.01	.34	-.18	.15	.71

*This is close to the "test" factor referred to as Factor VI after rotation.

Table A-2

Oblique Factor Loadings

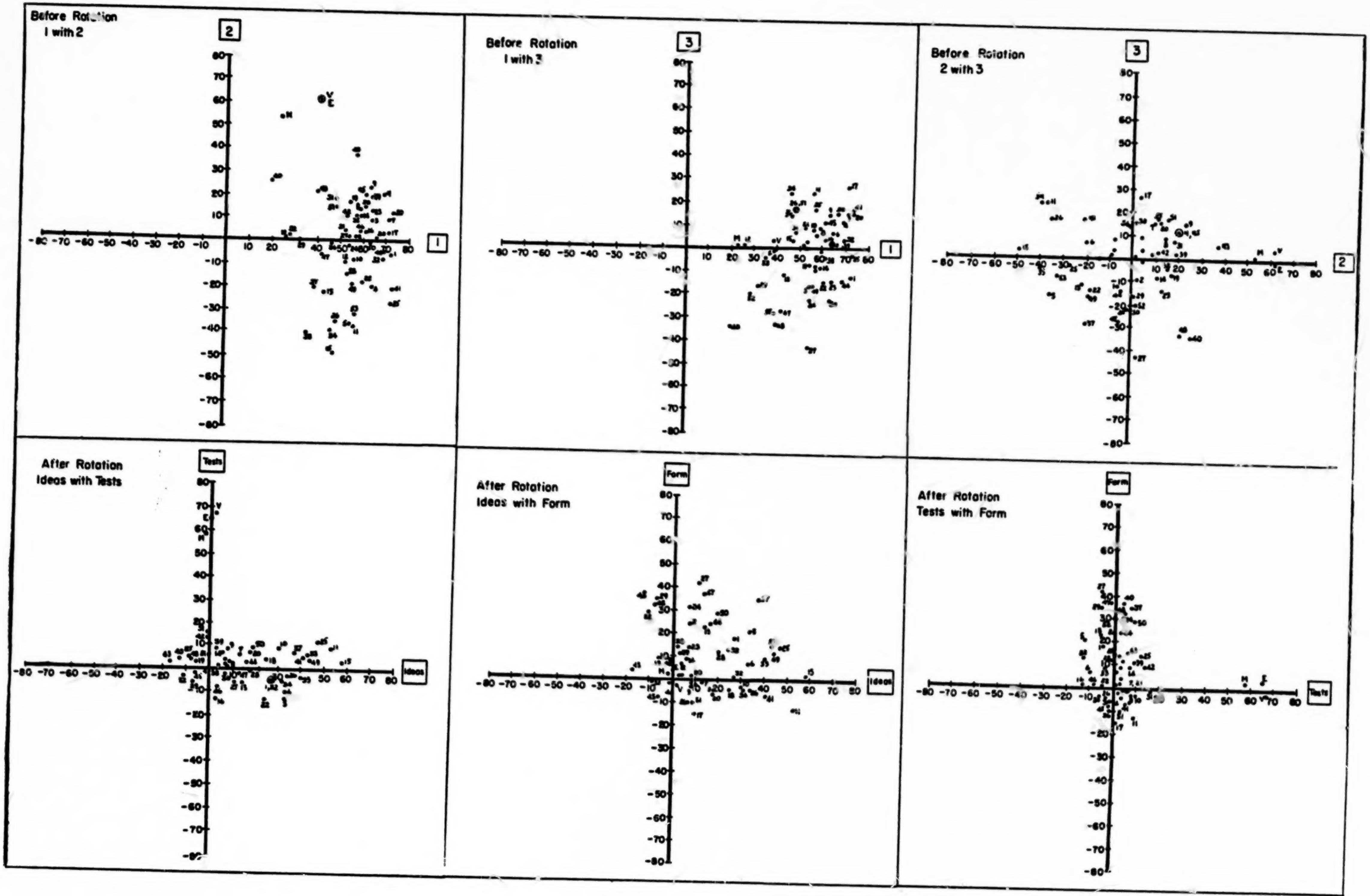
Reader	Factor					
	I	II	III	IV	V	VI
English Teachers						
1	.27	.18	-.02	.30	.05	-.05
2	.07	.24	.24	.14	.06	.00
3	.18	-.05	-.20	.32	.16	-.01
4	.00	-.05	.04	.31	.20	.00
5	.33	.21	.10	-.03	.17	-.13
6	.33	.07	.13	.35	-.11	-.10
7	.13	-.01	.15	.30	.06	.06
8	.10	-.02	-.00	.36	.02	.02
9	.08	-.03	.04	.47	-.09	.08
10	.30	-.03	.02	.15	.04	.09
Social Scientists						
11	.53	-.13	.19	.11	-.12	.09
12	.04	.07	.13	.09	-.01	.01
13	.14	.23	.26	-.09	.16	-.06
14	.06	.09	-.01	.05	.29	.07
15	.58	.02	.11	-.03	-.06	.03
16	.03	.01	.02	.26	.21	-.13
17	.10	-.15	.19	.23	.17	.01
18	.25	-.04	-.07	.07	.18	.04
19	-.06	.09	.03	.03	.35	.03
Natural Scientists						
20	.18	-.06	.02	.38	.03	.07
21	.00	-.01	.21	-.04	.30	.06
22	.24	.13	-.06	.09	.25	-.13
23	.07	.14	-.07	.13	.29	.02
24	.07	.31	.18	-.08	.24	.04
25	.47	.14	.18	-.01	.07	.12
26	.34	-.04	.27	.04	-.01	-.03
27	.11	.42	-.11	.07	.23	-.06
Writers and Editors						
28	.20	.12	.19	.06	.09	-.01
29	-.07	.35	.33	.03	.02	-.06
30	.08	.01	.28	.19	.11	.01
31	-.01	-.01	.12	.01	.22	.16
32	.27	.01	-.02	.29	.08	-.05
33	.40	.09	.10	-.06	.02	-.04
34	.31	-.04	.41	.06	-.07	-.06
35	-.09	-.01	.34	.03	.26	.05
36	.07	-.10	.07	.09	.20	-.02
37	.37	.35	.06	.03	-.11	.07
Lawyers						
38	.02	.14	.07	.26	.11	.00
39	.03	.11	.14	.24	.08	.09
40	-.13	.37	-.04	.07	.06	.04
41	.41	-.07	.11	.21	.01	.05
42	-.01	.09	.26	-.06	.25	.14
43	-.17	.04	.06	.33	.15	.03
44	.16	.24	.17	.01	.24	.03
45	-.06	-.08	.01	.22	.30	-.02
46	.04	.01	-.03	.14	.31	-.08
Business Executives						
47	.13	.37	.11	.07	.02	-.02
48	-.08	.32	-.06	-.00	.27	.04
49	.44	.12	-.17	.11	.06	.04
50	.19	.29	.08	.02	.17	.09
51	.09	-.10	-.02	.40	-.07	.03
52	-.11	.29	.23	-.13	.22	-.03
53	.43	.15	.15	-.07	.09	.06
Tests						
SAT-V	.02	-.02	-.04	.02	.01	.67
SAT-M	-.02	.03	.02	-.02	-.05	.58
ECT	.00	.04	-.03	.02	.01	.65

Table A-3

FACTOR INTERCORRELATIONS

		I	II	III	IV	V	VI
I	Ideas		.03	.27	.25	.55	-.02
II	Form	.03		-.03	.33	.10	.16
III	Flavor	.27	-.03		.45	.34	.07
IV	Mechanics	.25	.33	.45		.63	.50
V	Wording	.55	.10	.34	.63		.45
VI	Tests	-.02	.16	.07	.50	.45	

Figure A-1
Plots of Factor Loadings Before and After Rotation



Appendix B

Table of Intercorrelations

Appendix B

Intercorrelations of Readers (1-53) and Tests (V, M, ECT,

	Reader	Reader																										
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
English Teachers	1		.42	.51	.51	.46	.53	.47	.43	.47	.39	.39	.14	.29	.39	.39	.46	.48	.35	.29	.56	.33	.36	.49	.39	.53	.28	.42
	2	.42		.31	.44	.27	.39	.41	.37	.43	.30	.26	.14	.31	.28	.22	.38	.35	.19	.28	.43	.33	.36	.29	.34	.29	.29	.35
	3	.51	.31		.43	.33	.37	.49	.41	.43	.36	.31	.23	.18	.36	.24	.42	.46	.34	.38	.55	.31	.42	.39	.29	.41	.25	.39
	4	.51	.44	.43		.33	.40	.45	.51	.46	.39	.29	.17	.16	.42	.18	.46	.52	.33	.40	.54	.41	.38	.46	.35	.38	.30	.32
	5	.46	.27	.33	.33		.36	.37	.14	.23	.32	.39	.12	.35	.29	.47	.35	.35	.31	.24	.34	.28	.46	.31	.33	.49	.33	.37
	6	.53	.39	.37	.40	.36		.46	.41	.42	.34	.43	.18	.34	.32	.39	.38	.48	.27	.21	.48	.26	.38	.39	.28	.49	.48	.33
	7	.47	.41	.49	.45	.37	.46		.42	.53	.36	.44	.22	.26	.40	.27	.46	.58	.37	.38	.62	.37	.38	.45	.35	.45	.30	.34
	8	.43	.37	.41	.51	.14	.41	.42		.43	.30	.24	.16	.21	.36	.16	.38	.45	.25	.29	.47	.28	.29	.39	.32	.35	.30	.32
	9	.47	.43	.43	.46	.23	.42	.53	.43		.39	.35	.14	.25	.34	.13	.39	.50	.27	.33	.62	.26	.31	.31	.30	.34	.14	.31
	10	.39	.30	.36	.39	.32	.34	.36	.30	.39		.40	.16	.26	.36	.29	.32	.42	.35	33	.41	.27	.30	.40	.32	.39	.32	.24
Social Scientists	11	.39	.26	.31	.29	.39	.43	.44	.24	.35	.40		.14	.28	.32	.49	.25	.42	.43	.22	.40	.28	.37	.26	.27	.55	.40	.19
	12	.14	.14	.23	.17	.12	.18	.22	.16	.14	.16	.14		.16	.15	.06	.07	.14	.01	.21	.22	.15	.11	.18	.12	.19	.12	.09
	13	.29	.31	.18	.16	.35	.34	.26	.21	.25	.26	.28	.16		.24	.26	.25	.28	.16	.29	.23	.26	.37	.26	.35	.37	.26	.23
	14	.39	.28	.36	.42	.29	.32	.40	.36	.34	.36	.32	.15	.24		.22	.33	.42	.33	.38	.44	.33	.36	.47	.36	.46	.16	.36
	15	.39	.22	.24	.18	.47	.39	.27	.16	.13	.29	.49	.06	.26	.22		.28	.30	.34	.10	.29	.24	.36	.22	.23	.52	.43	.20
	16	.46	.38	.42	.46	.35	.38	.46	.38	.39	.32	.25	.07	.25	.33	.28		.48	.34	.34	.43	.32	.39	.43	.22	.41	.29	.32
	17	.48	.35	.46	.52	.35	.48	.58	.45	.50	.42	.42	.14	.28	.42	.30	.48		.40	.39	.53	.40	.37	.39	.31	.49	.43	.30
	18	.35	.19	.34	.33	.31	.27	.37	.25	.27	.35	.43	.01	.16	.33	.34	.34	.40		.33	.37	.30	.36	.32	.26	.43	.28	.33
	19	.29	.28	.38	.40	.24	.21	.38	.29	.33	.33	.22	.21	.29	.38	.10	.34	.39	.33		.40	.34	.32	.39	.33	.33	.18	.35
Natural Scientists	20	.56	.43	.55	.54	.34	.48	.62	.47	.62	.41	.40	.22	.23	.44	.29	.43	.53	.37	.40		.35	.40	.43	.32	.47	.29	.31
	21	.33	.33	.31	.41	.28	.26	.37	.28	.26	.27	.28	.15	.26	.33	.24	.32	.40	.30	.34	.35		.32	.33	.37	.34	.23	.24
	22	.36	.36	.42	.38	.46	.38	.38	.29	.31	.30	.37	.11	.37	.36	.36	.39	.37	.36	.32	.40	.32		.45	.32	.49	.27	.40
	23	.49	.29	.39	.46	.31	.39	.45	.39	.31	.40	.26	.18	.26	.47	.22	.43	.39	.32	.39	.43	.33	.45		.43	.45	.22	.37
	24	.39	.34	.29	.35	.33	.28	.35	.32	.30	.32	.27	.12	.35	.36	.23	.22	.31	.26	.33	.32	.37	.32	.43		.37	.21	.33
	25	.53	.29	.41	.38	.49	.49	.45	.35	.34	.39	.55	.19	.37	.46	.52	.41	.49	.43	.33	.47	.34	.49	.45	.37		.42	.39
	26	.28	.29	.25	.30	.33	.46	.30	.30	.14	.32	.40	.12	.26	.16	.43	.29	.43	.28	.18	.29	.23	.27	.22	.21	.42		.22
	27	.42	.35	.39	.32	.37	.33	.34	.32	.31	.24	.19	.09	.23	.36	.20	.32	.30	.33	.35	.31	.24	.40	.37	.33	.39	.22	
Writers & Editors	28	.43	.35	.33	.26	.26	.36	.36	.30	.29	.35	.34	.17	.20	.29	.29	.32	.36	.23	.32	.36	.26	.27	.32	.34	.44	.32	.26
	29	.28	.30	.08	.18	.24	.23	.28	.15	.24	.08	.15	.13	.23	.14	.05	.19	.22	.08	.16	.22	.17	.19	.18	.26	.17	.13	.24
	30	.42	.38	.35	.54	.31	.47	.49	.47	.40	.27	.44	.24	.28	.36	.29	.39	.51	.30	.32	.47	.41	.41	.40	.32	.52	.33	.29
	31	.25	.19	.36	.31	.21	.24	.38	.33	.32	.24	.21	.10	.11	.36	.18	.29	.44	.30	.30	.39	.32	.21	.26	.24	.31	.19	.23
	32	.55	.33	.49	.44	.39	.50	.51	.45	.44	.40	.45	.15	.35	.41	.29	.42	.48	.39	.35	.53	.29	.46	.54	.34	.51	.34	.30
	33	.30	.22	.15	.15	.32	.26	.19	.20	.11	.24	.32	.07	.24	.12	.42	.19	.15	.20	.08	.19	.21	.30	.20	.20	.43	.23	.17
	34	.30	.24	.20	.27	.41	.43	.35	.23	.24	.27	.49	.21	.27	.14	.39	.27	.43	.24	.09	.33	.26	.25	.22	.26	.48	.42	.13
	35	.34	.39	.32	.42	.20	.27	.42	.32	.41	.32	.28	.14	.31	.38	.16	.36	.52	.25	.33	.40	.36	.23	.37	.29	.37	.25	.21
	36	.29	.22	.33	.39	.26	.27	.31	.24	.26	.28	.28	.12	.16	.29	.23	.24	.37	.20	.25	.33	.20	.25	.29	.23	.33	.26	.15
	37	.37	.28	.24	.19	.27	.31	.22	.20	.18	.25	.21	.10	.18	.15	.37	.17	.12	.20	.17	.22	.11	.22	.21	.25	.38	.23	.30
Lawyers	38	.50	.34	.39	.46	.27	.37	.44	.39	.47	.25	.24	.18	.15	.34	.20	.42	.42	.20	.38	.47	.33	.31	.39	.30	.43	.24	.31
	39	.47	.44	.41	.52	.31	.40	.51	.38	.54	.32	.26	.23	.24	.35	.23	.41	.44	.33	.33	.52	.35	.32	.37	.31	.40	.22	.35
	40	.17	.14	.12	.14	.00	.11	.14	.19	.13	.08	.05	.14	.04	.16	-.04	.15	.06	.13	.15	.10	-.01	.06	.24	.17	.10	-.06	.33
	41	.51	.44	.45	.51	.43	.47	.50	.43	.42	.48	.53	.15	.31	.39	.51	.43	.52	.37	.37	.49	.37	.46	.37	.29	.62	.46	.33
	42	.31	.33	.25	.39	.19	.30	.40	.30	.32	.23	.27	.10	.26	.33	.22	.27	.37	.27	.36	.39	.35	.23	.34	.33	.35	.28	.25
	43	.39	.34	.34	.55	.15	.40	.46	.40	.44	.22	.12	.18	.14	.32	.07	.44	.47	.21	.37	.42	.30	.27	.42	.28	.31	.15	.32
	44	.51	.39	.42	.44	.43	.39	.48	.39	.40	.38	.36	.12	.34	.43	.34	.41	.48	.41	.42	.47	.40	.40	.39	.38	.53	.39	.49
	45	.45	.38	.42	.52	.22	.34	.43	.36	.44	.35	.23	.16	.15	.31	.16	.39	.43	.34	.35	.46	.39	.37	.50	.31	.30	.21	.29
	46	.42	.23	.45	.44	.34	.33	.42	.37	.30	.26	.28	.13	.22	.34	.26	.44	.44	.31	.36	.45	.30	.39	.45	.33	.38	.30	.38
Business Executives	47	.31	.29	.16	.26	.28	.30	.30	.25	.25	.23	.25	.14	.18	.26	.13	.25	.21	.17	.24	.25	.15	.27	.32	.31	.29	.19	.36
	48	.34	.28	.26	.29	.21	.16	.21	.21	.28	.16	.03	.09	.15	.27	.09	.20	.22	.18	.23	.26	.16	.25	.27	.26	.21	.03	.37
	49	.42	.26	.41	.32	.39	.40	.32	.31	.30	.32	.40	.16	.29	.35	.38	.27	.35	.40	.32	.38	.23	.45	.40	.34	.48	.27	.45
	50	.49	.42	.41	.40	.38	.42	.40	.33	.32	.36	.28	.13	.29	.45	.34	.34	.40	26	.26	.45	.36	.42	.41	.39	.47	.30	.42
	51	.31	.25	.30	.39	.13	.36	.40	.35	.48	.35	.28	.09	.17	.29	.19	.31	.35	.25	.27	.42	.22	.24	.30	.21	.25	.19	.21
	52	.18	.19	.16	.14	.20	.15	.16	.04	.12	.10	.03	.03	.19	.17	.05	.21	.16	.12	.18	.19	.07	.14	.16	.19	.23	.11	.15
	53	.40	.34	.33	.33	.43	.36	.31	.27	.23	.40	.48	.19	.30	.27	.48	.30	.35	.31	.28	.37	.22	.44	.34	.33	.57	.42	.32
Test	V	.21	.17	.31	.34	.01	.09	.32	.28	.33	.18	.08	.08	.07	.32	-.01	.19	.29	.15	.29	.34	.28	.13	.35	.19	.22	-.03	.15
	M	.12	.14	.13	.22	-.05	-.03	.20	.19	.26	.15	.04	.06	-.03	.16	-.10	.08	.14	.10	.17	.22	.09	.04	.20	.13	.11	-.03	.11
	ECT	.22	.21	.32	.34	.00	.11	.32	.26	.32	.19	.05	.06	.02	.26	.00	.23	.27	.19	.27	.33	.24	.08	.35	.18	.21	.05	.22

Reader																										Test		
28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	V	M	ECT
.43	.28	.42	.25	.55	.30	.30	.34	.29	.37	.50	.47	.17	.51	.31	.39	.51	.45	.42	.31	.34	.42	.49	.31	.18	.40	.21	.12	.22
.35	.30	.38	.19	.33	.22	.24	.39	.22	.28	.34	.44	.14	.44	.33	.34	.39	.38	.23	.29	.28	.26	.42	.25	.19	.34	.17	.14	.21
.33	.08	.35	.36	.49	.15	.20	.32	.33	.24	.39	.41	.12	.45	.25	.34	.42	.42	.45	.16	.26	.41	.41	.30	.16	.33	.31	.13	.32
.26	.18	.54	.31	.44	.15	.27	.42	.39	.19	.46	.52	.14	.51	.39	.55	.44	.52	.44	.26	.29	.32	.40	.39	.14	.33	.34	.22	.34
.26	.24	.31	.21	.39	.32	.41	.20	.26	.27	.27	.31	.00	.43	.19	.15	.43	.22	.34	.28	.21	.39	.38	.13	.20	.43	.01	-.05	.00
.36	.23	.47	.24	.50	.26	.43	.27	.27	.31	.37	.40	.11	.47	.30	.40	.39	.34	.33	.30	.16	.40	.42	.36	.15	.36	.09	-.03	.11
.36	.28	.49	.38	.51	.19	.35	.42	.31	.22	.44	.51	.14	.50	.40	.46	.48	.43	.42	.30	.21	.32	.40	.40	.16	.31	.32	.20	.32
.30	.15	.47	.33	.45	.20	.23	.32	.24	.20	.39	.38	.19	.43	.30	.40	.39	.36	.37	.25	.21	.31	.33	.35	.04	.27	.28	.19	.26
.29	.24	.40	.32	.44	.11	.24	.41	.26	.18	.47	.54	.13	.42	.32	.44	.40	.44	.30	.25	.28	.30	.32	.48	.12	.23	.33	.26	.32
.35	.08	.27	.24	.40	.24	.27	.32	.28	.25	.25	.32	.08	.48	.23	.22	.38	.35	.26	.23	.16	.32	.36	.35	.10	.40	.18	.15	.19
.34	.15	.44	.21	.45	.32	.49	.28	.28	.21	.24	.26	-.05	.53	.27	.12	.36	.23	.28	.25	.03	.40	.28	.28	.03	.48	.08	.04	.05
.17	.13	.24	.10	.15	.07	.21	.14	.12	.10	.18	.23	.14	.15	.10	.18	.12	.16	.13	.14	.09	.16	.13	.09	.03	.19	.08	.06	.06
.20	.23	.28	.11	.35	.24	.27	.31	.16	.18	.15	.24	.04	.31	.26	.14	.34	.15	.22	.18	.15	.29	.29	.17	.19	.30	.07	-.03	.02
.29	.14	.36	.36	.41	.12	.14	.38	.29	.15	.34	.35	.16	.39	.33	.32	.43	.31	.34	.26	.27	.35	.45	.29	.17	.27	.32	.16	.26
.29	.05	.29	.18	.29	.42	.39	.16	.23	.37	.20	.23	-.04	.51	.22	.07	.34	.16	.26	.13	.09	.38	.34	.19	.05	.48	-.01	-.10	.00
.32	.19	.39	.29	.42	.19	.27	.36	.24	.17	.42	.41	.15	.43	.27	.44	.41	.39	.44	.25	.20	.27	.34	.31	.21	.30	.19	.08	.23
.36	.22	.51	.44	.48	.15	.43	.52	.37	.12	.42	.44	.06	.52	.37	.47	.48	.43	.44	.21	.22	.35	.40	.35	.16	.35	.29	.14	.27
.23	.08	.30	.30	.39	.20	.24	.25	.20	.20	.20	.33	.13	.37	.27	.21	.41	.34	.31	.17	.18	.40	.26	.25	.12	.31	.15	.10	.19
.32	.16	.32	.30	.35	.08	.09	.33	.25	.17	.38	.33	.15	.37	.36	.37	.42	.35	.36	.24	.23	.32	.26	.27	.18	.28	.29	.17	.27
.36	.22	.47	.39	.53	.19	.33	.40	.33	.22	.47	.52	.10	.49	.39	.42	.47	.46	.45	.25	.26	.38	.45	.42	.19	.37	.34	.22	.33
.26	.17	.41	.32	.29	.21	.26	.36	.20	.11	.33	.35	-.01	.37	.35	.30	.40	.39	.30	.15	.16	.23	.36	.22	.07	.22	.28	.09	.24
.27	.19	.41	.21	.46	.30	.25	.23	.25	.22	.31	.32	.06	.46	.23	.27	.40	.37	.39	.27	.25	.45	.42	.24	.14	.44	.13	.04	.08
.32	.18	.40	.26	.54	.20	.22	.37	.29	.21	.39	.37	.24	.37	.34	.42	.39	.50	.45	.32	.27	.40	.41	.30	.16	.34	.35	.20	.35
.34	.26	.32	.24	.34	.20	.26	.29	.23	.25	.30	.31	.17	.29	.33	.28	.38	.31	.33	.31	.26	.34	.39	.21	.19	.33	.19	.13	.18
.44	.17	52	.31	.51	.43	.48	.37	.33	.38	.43	.40	.10	.62	.35	.31	.53	.30	.38	.29	.21	.48	.47	.25	.23	.57	.22	.11	.21
.32	.13	.33	.19	.34	.23	.43	.25	.26	.23	.24	.22	-.06	.46	.28	.15	.39	.21	.30	.19	.03	.27	.30	.19	.11	.42	-.03	-.03	.05
.26	.24	.29	.23	.30	.17	.13	.21	.15	.30	.31	.35	.33	.33	.25	.32	.49	.29	.38	.36	.37	.45	.42	.21	.15	.32	.15	.11	.22
	.19	.34	.26	.38	.23	.30	.29	.22	.22	.26	.33	.02	.44	.33	.27	.38	.31	.25	.29	.24	.28	.33	.19	.11	.39	.11	.00	.11
.19		.29	.16	.16	.07	.21	.23	.11	.10	.21	.23	.20	.15	.08	.22	.24	.10	.11	.33	.13	.05	.20	.18	.20	.19	.05	.02	.10
.34	.29		.26	.40	.24	.34	.39	.30	.17	.44	.45	[illegible]	.50	.36	.51	.46	.46	.41	.27	.15	.27	.35	.34	.22	.40	.24	.16	.22
.26	.16	.26		.28	.09	.19	.36	.27	.13	.23	.29	.13	.30	.25	.27	.33	.33	.25	.19	.28	.17	.31	.25	.19	.22	.34	.18	.31
.38	.16	.40	.28		.28	.36	.34	.32	.32	.45	.37	.03	.51	.34	.36	.38	.44	.34	.30	.28	.37	.44	.35	.20	.36	.22	.10	.19
.23	.07	.24	.09	.28		.29	.17	.16	.26	.12	.11	-.03	.40	.14	-.01	.30	.20	.23	.18	.03	.33	.19	.03	.10	.30	-.05	-.03	-.10
.30	.21	.34	.19	.36	.29		.35	.25	.25	.21	.23	-.01	.37	.25	.18	.29	.19	.23	.19	-.02	.24	.22	.16	.13	.32	-.04	-.07	-.08
.29	.23	.39	.36	.34	.17	.35		.31	.11	.32	.39	.14	.39	.35	.37	.42	.40	.35	.18	.20	.20	.31	.30	.18	.28	.25	.18	.22
.22	.11	.30	.27	.32	.16	.25	.31		.09	.25	.30	.04	.31	.25	.29	.25	.34	.29	.09	.22	.31	.27	.21	.15	.29	.14	.09	.15
.22	.10	.17	.13	.32	.26	.25	.11	.09		.31	.19	.13	.34	.24	.17	.29	.13	.14	.29	.21	.30	.35	.13	.19	.30	.09	-.01	.10
.26	.21	.44	.23	.45	.12	.21	.32	.25	.31		.37	.18	.44	.34	.47	.48	.38	.41	.24	.28	.22	.39	.31	.28	.26	.31	.16	.27
.33	.23	.45	.29	.37	.11	.23	.39	.30	.19	.37		.19	.46	.39	.47	.41	.42	.32	.29	.30	.30	.42	.31	.23	.25	.28	.26	.37
.02	.20	.10	.13	.03	-.03	-.01	.14	.04	.13	.18	.19		.00	.09	.25	.20	.06	.13	.19	.20	.09	.14	.16	.14	-.02	.19	.16	.17
.44	.15	.50	.30	.51	.40	.37	.39	.31	.34	.44	.46	.00		.40	.36	.49	.44	.38	.24	.17	.43	.42	.37	.08	.50	.22	.06	.18
.33	.08	.36	.25	.34	.14	.25	.35	.25	.24	.34	.39	.09	.40		.34	.41	.42	.37	.15	.29	.18	.38	.22	.27	.25	.25	.23	.29
.27	.22	.51	.27	.36	-.01	.18	.37	.29	.17	.47	.47	.25	.36	.34		.34	.47	.36	.17	.32	.20	.31	.35	.14	.14	.38	.28	.39
.38	.24	.46	.33	.38	.30	.29	.42	.25	.29	.48	.41	.20	.49	.41	.34		.38	.39	.31	.30	.38	.47	.30	.26	.44	.24	.10	.26
.31	.10	.46	.33	.44	.20	.19	.40	.34	.13	.38	.42	.06	.44	.42	.47	.38		.37	.19	.33	.25	.31	.31	.15	.29	.29	.23	.28
.25	.11	.41	.25	.34	.23	.23	.35	.29	.14	.41	.32	.13	.38	.37	.36	.39	.37		.30	.23	.30	.31	.26	.13	.33	.23	.11	.25
.29	.33	.27	.19	.30	.18	.19	.18	.09	.29	.24	.29	.19	24	.15	.17	.31	.19	.30		.20	.26	.23	.18	.12	.28	.09	.10	.11
.24	.13	.15	.28	.28	.03	-.02	.20	.22	.21	.28	.30	.20	.17	.29	.32	.30	.33	.23	.20		.18	.36	.16	.20	.26	.24	.17	26
.28	.05	.27	.17	.37	.33	.24	.20	.31	.30	.22	.30	.09	.43	.18	.20	.38	.25	.30	.26	.18		.38	.22	.07	.42	.17	.12	.14
.33	.20	.35	.31	.44	.19	.22	.31	.27	.35	.39	.42	.14	.42	.38	.31	.47	.31	.31	.23	.36	.38		.23	.25	.34	.26	.15	.27
.19	.18	.34	.25	.35	.03	.16	.30	.21	.13	.31	.31	.16	.37	.22	.35	.30	.31	.26	.18	.16	.22	.23		.12	.21	.26	.11	.26
.11	.20	.22	.19	.20	.10	.13	.18	.15	.19	.28	.23	.14	.08	.27	.14	.26	.15	.13	.12	.20	.07	.25	.12		.18	.03	.07	.09
.39	.19	.40	.22	.36	.30	.32	.28	.29	.30	.26	.25	-.02	.50	.25	.14	.44	.29	.33	.28	.26	.42	.34	.21	.18		.08	.04	.13
.11	.05	.24	.34	.22	-.05	-.04	.25	.14	.09	.31	.28	.19	.22	.25	.38	.24	.29	.23	.09	.24	.17	.26	.26	.03	.08		.57	.76
.00	.02	.16	.18	.10	-.03	-.07	.18	.09	-.01	.16	.26	.16	.06	.23	.28	.10	.23	.11	.10	.17	.12	.15	.11	.07	.04	.57		.53
.11	.10	.22	.31	.19	-.10	-.08	.22	.15	.10	.27	.37	.17	.18	.29	.39	.26	.28	.25	.11	.26	.14	.27	.26	.09	.13	.76	.53	

Appendix C

THEME TOPICS

1.

Here are two points of view on the question, "Who should go to college?"

(a) Students who score low on entrance tests often gain the most from college. America is known as a land based on the ideals of freedom and equality. If this is so, higher education should be available to everyone, not just a chosen few.

(b) I have difficulty in subscribing to the theory, apparently regarded as inherent in one version of democracy, that every child of a taxpayer has a right to enter a tax-supported institution of higher education. In my view, a university is a place where the more mature or more promising intellects can pursue intellectual studies at a high level for no other reason than from a spirit of compelling intellectual curiosity.

College enrollments have already begun to respond to prosperity and to higher birth rates, and it is estimated that by 1970 most colleges must either double both plant and staff or resort to a policy of selective admission. The problem of coping with an unprecedented increase in enrollment is exciting lively attention and deep concern. The solution to the many problems arising out of this situation depends to a large extent on our answer to the question, "Who should go to college?" Discuss and defend your ideas on who should receive higher education in a period in which it will be extremely difficult to accommodate all who apply for admission to college.

2.

Teen-agers often find their situation confusing, because they are sometimes treated as adults, sometimes as children. When should teen-agers be treated as adults? You might consider some of the following questions in making your decision:

(a) Are you mature when you reach a certain age? (The government considers eighteen a satisfactory age for military service.)

(b) Are you mature when you graduate from high school? (The high school diploma indicates that you are ready for work or for college.)

(c) Are you mature when you have a regular job and are self-supporting? (Many high school students have regular jobs.)

(d) Are you mature when you can accept responsibility? (Some high school freshmen are more conscientious and dependable than seniors.)

BLANK PAGE

When do you think teen-agers should be treated as adults? Give reasons for your opinion and support it by specific suggestions or examples.

Appendix D

Readers

English Teachers

Mark Ashin	Department of English	University of Chicago
Douglas Beers	Department of English	Middlebury College
Richard Bozorth	Department of English	University of Pennsylvania
Clarence Derrick	Department of English	University of Florida
Ralph C. Leyden	Department of English	Stephens College
C. F. Main	Department of English	Rutgers University
David Novarr	Department of English	Cornell University
Osmond Palmer	Department of English	Michigan State University
Macklin Thomas	Department of English	Chicago City Junior College
R. L. Zetler	Department of English	Chatham College

Social Scientists

Henry Bragdon	Department of History	Phillips Exeter Academy
Henry Graff	Department of History	Columbia University
Preston E. James	Department of Geography	Syracuse University
Elihu Katz	Department of Sociology	University of Chicago
Norman D. Palmer	Dept. of Political Science	University of Pennsylvania
Harold Stein	Department of Government	Princeton University
Joseph R. Strayer	Department of History	Princeton University
Melvin Tumin	Department of Sociology	Princeton University
Stanton Wheeler	Department of Sociology	Harvard University

Natural Scientists

	Peter Axel	Department of Physics	University of Illinois
(Miss)	Janet B. Guernsey	Department of Physics	Wellesley College
	Sheldon Judson	Department of Geology	Princeton University
	A. Kerman	Department of Physics	Massachusetts Institute of Technology
	John R. Pruett	Department of Physics	Bryn Mawr College
	Wendell H. Taylor	Department of Science	Lawrenceville School
	Robert E. Todd	Department of Zoology	Colgate University
	Frank Verbrugge	School of Physics	University of Minnesota

Writers & Editors

	William B. Arthur	Look Magazine
	Harold Baron	Free lance
	John L. Cobbs	Business Week
(Miss)	Michael Drury	Free lance
	David R. Ebbitt	Scott, Foresman & Co.
(Mrs.)	Amelia Lobsenz	Lobsenz & Wirsig, Inc.
	Llewellyn Miller	Free lance
	C. B. Roberts	This Week Magazine
(Mrs.)	Beatrice Schapper	Free lance
	Woodrow Wirsig	Printers' Ink

Lawyers

Vaughn C. Ball	College of Law, Ohio State University
Monroe H. Freedman	The Law School, George Washington University
Gerald Gunther	School of Law, Columbia University
Saul Mendlovitz	School of Law, Rutgers University
Robert S. Pasley	Cornell Law School
Melvin G. Shimm	School of Law, Duke University
Emerson G. Spies	University of Virginia Law School
Saul Touster	University of Buffalo Law School
William F. Young	School of Law, Columbia University

Business Executives

Anton W. Bondy	Lever Brothers Co.
Elric G. Endersby	Old Deerfield, Inc.
Peter Lauck, III	Stroheim & Romann, Inc.
John H. Platten, Jr.	J. A. Ward, Inc.
Gene Shalit	Barkas & Shalit, Inc.
Ralph J. Sharp	Street & Smith Publications, Inc.
William L. Wilson, Jr.	Commercial Investment Trust, Inc.

Appendix E

Correlations of Readers and Tests with Consensus Scores

Reader	Correlation	Reader	Correlation
English Teachers		**Writers and Editors**	
1	70	28	54
2	57	29	36
3	62	30	66
4	67	31	46
5	55	32	68
6	63	33	39
7	69	34	49
8	56	35	56
9	61	36	48
10	58	37	42
Social Scientists		**Lawyers**	
11	59	38	60
12	28	39	63
13	45	40	22
14	58	41	72
15	48	42	54
16	58	43	56
17	70	44	70
18	52	45	58
19	53	46	56
Natural Scientists		**Business Executives**	
20	71	47	44
21	52	48	41
22	60	49	54
23	59	50	62
24	55	51	48
25	73	52	31
26	49	53	58
27	54		
		Tests	
		SAT-V	34
		SAT-M	19
		ECT	33

Appendix F

Distribution of Grades for Each Reader

Reader	Grade								
	1	2	3	4	5	6	7	8	9
English Teachers									
1	13	25	36	45	57	47	39	22	12
2	20	31	35	32	45	48	43	23	19
3	13	18	48	43	61	48	37	16	12
4	12	41	46	44	46	28	37	30	12
5	28	18	20	44	79	51	23	21	12
6	11	13	37	66	68	48	25	16	12
7	12	21	36	54	61	55	26	19	12
8	15	22	40	58	74	33	27	17	10
9	15	26	35	47	59	43	34	23	14
10	49	34	33	42	42	29	27	17	23
Social Scientists									
11	14	33	71	52	50	26	25	13	12
12	23	36	43	30	20	50	40	35	19
13	36	39	41	57	70	14	14	12	13
14	10	21	64	76	42	43	15	13	12
15	14	27	29	44	59	50	37	22	14
16	14	30	59	62	47	37	22	13	12
17	12	18	31	57	59	48	37	24	10
18	18	37	42	58	59	30	28	12	12
19	13	17	30	46	87	45	30	16	12
Natural Scientists									
20	17	35	44	45	48	40	27	24	16
21	15	31	36	38	47	38	43	32	16
22	15	17	39	49	66	42	35	19	14
23	26	38	50	63	39	27	26	12	15
24	12	19	37	52	57	45	41	23	10
25	13	22	31	40	60	53	41	22	14
26	7	20	33	49	86	48	29	16	8
27	21	34	47	56	58	28	23	15	14
Writers and Editors									
28	15	30	48	41	53	41	38	17	13
29	60	47	30	34	43	19	24	19	20
30	23	29	30	37	52	51	34	25	15
31	15	20	59	74	50	27	13	24	14
32	18	54	58	52	43	23	21	14	13
33	23	40	43	49	40	34	28	23	16
34	34	44	38	28	57	34	25	20	16
35	13	17	30	39	68	46	47	23	13
36	40	47	38	33	29	28	35	34	12
37	14	28	56	65	39	41	27	13	13

Reader	Grade								
	1	2	3	4	5	6	7	8	9
Lawyers									
38	13	17	40	51	62	49	34	17	13
39	16	20	33	46	55	50	38	24	14
40	17	28	27	47	67	40	32	18	20
41	12	31	35	55	63	40	33	15	12
42	12	31	80	89	34	14	12	12	12
43	14	16	39	49	54	47	40	25	12
44	21	23	37	57	55	39	28	21	15
45	14	26	35	46	56	51	37	17	14
46	12	20	21	62	94	48	15	12	12
Business Executives									
47	11	23	35	49	60	45	37	24	12
48	22	18	43	54	40	39	23	28	29
49	11	13	18	21	33	81	64	40	15
50	12	23	38	55	69	42	34	14	9
51	25	21	51	60	54	34	21	16	14
52	17	29	32	47	73	30	32	22	14
53	18	24	46	44	51	35	41	23	14
Mean	18	27	40	50	55	40	31	20	14
Range	7-60	13-54	18-80	21-89	20-94	14-81	12-64	12-40	8-29

Appendix G

Distribution of Grades for Each Student

Student	Grade								
	1	2	3	4	5	6	7	8	9
1	4	8	5	13	10	4	6	1	2
2	1		2		8	14	10	8	10
3		1	2	1	11	12	9	6	11
4		6	6	12	13	8	4	3	1
5	3	3	11	7	9	11	3	3	3
6	10	13	16	4	7	2	1		
7	1	2	3	5	6	6	8	7	15
8	2		10	12	16	5	5	3	
9	1	4	6	9	7	11	8	5	2
10	4	5	5	5	4	4	10	6	10
11		2	7	11	17	6	4	4	2
12	1	5	3	3	7	9	11	7	7
13	3	6	7	6	12	7	7	2	3
14	44	5	2		1			1	
15	9	13	9	8	5	5	4		
16	6	8	6	14	7	4	4	2	2
17	12	10	8	11	7	3	1		1
18	16	17	4	12	2	1			1
19	3	21	15	9	2	2	1		
20		2	4	4	11	11	9	6	6
21	2	10	12	11	7	7	2	2	
22	8	3	7	6	11	5	6	5	2
23	4	5	5	4	10	12	5	4	4
24		4	4	7	15	10	8	3	2
25	7	7	7	9	9	8	2	4	
26	1	1	2	4	7	4	10	13	11
27	8	3	5	10	16	3	4	3	1
28		6	10	12	12	6	6	1	
29	11	12	10	10	7	2	1		
30	2	4	3	8	9	5	11	5	6
31	4	5	8	10	9	8	5	2	2
32		3	5	5	15	11	7	5	2
33	1	4	9	7	14	10	7	1	
34	1	1	3	5	13	11	10	3	6
35	6	3	6	9	14	7	4	2	2
36		6	6	8	11	5	5	9	3
37	4	11	12	9	4	5	2	4	2
38	2	1	5	10	14	14	4	2	1
39	6	7	15	12	12	1			
40	2	1	6	5	8	9	14	6	2

Student	Grade								
	1	2	3	4	5	6	7	8	9
41	1	1	6	5	11	13	3	6	7
42		3	3	9	14	5	9	7	3
43	7	7	11	3	8	5	3	5	4
44	5	3	8	11	6	7	6	4	3
45	2	3	7	12	12	10	3	4	
46	1	4	7	8	14	6	8	4	1
47	5	4	8	9	11	7	5		4
48	2	1	6	6	20	8	7	3	
49	4	3	6	9	12	7	4	6	2
50		1	1	1	10	7	11	8	14
51	13	9	8	11	7	3	2		
52	2	1	10	12	10	11	2	4	2
53	2	3	1	8	13	16	2	5	3
54	1	4	10	10	14	9	2	2	1
55		2	7	12	13	5	10	1	3
56	1	8	12	17	6	2	3	2	2
57	1	4	6	13	5	13	4	4	3
58	1	3	5	10	13	10	3	6	2
59	1		5	10	13	3	14	6	1
60		2	3	2	2	8	10	13	13
61	3	4	12	12	15	5	1	1	
62	5	20	6	12	4	3	3		
63			1	11	14	6	9	5	7
64	11	15	7	10	5	4	1		
65	1	1	2	9	15	6	7	9	3
66	1		3	4	8	9	10	5	13
67	11	5	8	9	7	3	5	1	4
68	2	1	11	7	17	5	8	1	1
69	2	5	8	6	5	7	8	8	4
70	2	1	6	10	13	10	9	1	1
71	6	14	3	5	4	7	7	5	2
72		1	6	13	14	7	4	4	4
73		1	2	6	8	13	8	10	5
74	1	5	5	7	10	9	9	5	2
75	4	9	16	13	8	3			
76	5	3	12	11	13	5	2	2	
77	2	12	9	10	9	3	6	1	1
78	1		8	5	11	10	10	3	5
79	3	4	8	9	17	7		4	1
80	3	8	13	10	8	7	2	2	
81	1		1	2	4	8	14	12	11
82		4	6	9	14	8	7	3	2
83	6	15	13	11	7	1			
84	16	7	10	9	3	3	2	3	
85	2	10	10	13	6	7	2	3	
86		1	1	1	2	6	9	12	21
88	8	18	17	6	3	1			
89	1		6	10	8	11	9	5	3
90	13	15	13	4	3	4		1	

Student	Grade								
	1	2	3	4	5	6	7	8	9
91				4	6	6	9	7	21
92	1	2	2	7	11	10	9	7	4
93	3	2	10	14	6	9	6	3	
94			2	3	12	12	15	4	5
95	4	4	9	8	18	7	3		
96	6	17	17	7	6				
97	4	9	16	11	8	5			
98		1		1	9	5	8	15	14
99	2	2		7	5	10	18	5	4
100	2	5	12	13	10	8	2	1	
101	8	9	9	8	9	4	5	1	
102	9	13	12	12	4	2	1		
103	16	18	10	6	1		1	1	
104	1	1	7	14	6	11	8	4	1
105	1	2	8	16	15	8	3		
106	3	10	10	13	11	5	1		
107	10	14	11	11	4	2		1	
108	11	9	16	6	7	1	2	1	
109	7	4	1	7	9	9	9	7	
110	1		2	2	10	19	6	9	4
111	3	2	1	11	21	5	5	3	2
112		2	5	7	12	12	8	3	4
113	2	6	8	8	9	6	7	4	3
114	3	8	17	16	2	4	2	1	
115	1	3	8	10	11	9	6	2	3
116		3	5	19	18	4	3	1	
117	1	6	6	11	14	9	2	1	3
118				7	16	18	6	4	2
119		2	6	8	14	10	10	2	1
120	1	5	9	10	14	8	4	2	
121			2	9	15	8	12	6	1
122		1	2	4	14	10	9	11	2
123	5	3	5	12	11	12	4		1
124	38	8	3	2	1	1			
125		1	6	9	12	10	10	2	3
126	1	3	13	14	15	6	1		
127	7	13	15	8	5	1	2	1	1
128			4	9	13	11	8	6	2
129	1	3	10	6	22	3	5	2	1
130	3	2	11	7	16	4	7	3	
131		3	5	6	11	9	10	4	5
132	6	10	12	10	8	2	4	1	
133	2	2	13	5	9	9	8	3	2
134		1	4	5	8	10	14	7	4
135	4	1	6	8	5	4	7	11	7
136	1	2	4	3	13	9	10	6	5
137	2	4	7	8	11	9	8	3	1
138	26	13	4	5	1	3	1		
139				10	15	14	7	4	3
140	10	10	14	10	7	1		1	

Student	Grade								
	1	2	3	4	5	6	7	8	9
141	1	7	9	13	11	9	1	2	
142	1	11	10	14	11	2	2	1	1
143			4	5	11	9	14	5	5
144	2	2	9	12	21	6		1	
145	2	2	6	4	2	10	14	4	9
146			2	9	16	10	7	7	2
147		4	6	19	14	8	2		
148	1	5	11	17	10	5	4		
149	1		9	16	20	6	1		
150	3	5	8	16	9	4	8		
151	4	7	7	9	13	9	2		2
152	3	5	5	14	14	7	4	1	
153	3	9	5	6	19	5	3	2	1
154	8	9	12	10	8	4	2		
155	3	3	6	7	16	5	8	3	2
156		3	4	7	13	8	10	4	4
157	6	8	12	11	8	6	1		1
158	1	2	1	1	5	2	8	12	21
159		4	5	9	12	13	8	2	
160	1		2	2	8	15	9	11	5
161		1	5	14	14	10	4	3	2
162	11	10	7	8	11	2	2	1	1
163	4	8	9	12	11	3	5	1	
164		2	1	9	13	9	8	6	5
165	3		9	15	18	5	1	1	1
166	12	13	13	5	4	2	1	3	
167	1		1	2	6	9	14	12	8
168	1		1	6	2	12	12	8	11
169	9	8	7	16	4	1	5	1	2
170	1	2	2	9	9	11	7	3	9
171	1	3	4	12	12	9	6	4	2
172	4	5	8	5	11	7	4	6	3
173	3	8	9	16	9	5	2		1
174			4	4	10	10	13	6	6
175	1		7	3	8	8	10	12	4
176	2	6	1	3	10	5	8	5	11
177	1		3	4	10	12	8	8	7
178	2	7	9	8	7	6	9	3	2
179	3	11	14	8	9	5	1	2	
180	1	1	6	4	14	6	5	8	8
181	1	1	1	4	9	15	12	7	3
182	3	2	10	9	7	5	7	6	4
183	1	3	15	8	20	5			1
184	4	14	12	9	6	6	2		
185	2	3	5	11	11	9	7	2	3
186		1	4	5	16	4	8	8	7
187	6	8	6	12	13	5	2	1	
188	6	6	20	4	11	2	1	1	2
189			5	6	12	12	10	4	4
190	1		1	2	1	2	6	12	28

Student	Grade								
	1	2	3	4	5	6	7	8	9
191	9	6	6	9	8	8	4	3	
192	4	7	14	9	13	2	1	3	
193		5	6	10	18	4	9		1
194		4	3	11	9	9	13	1	3
195	8	16	12	9	5	1		1	1
196	1	7	3	10	13	7	5	3	4
197	2	4	9	10	17	3	6	2	
198	1	4	12	7	13	8	5	1	2
199		3	6	7	10	9	10	5	3
200	3	4	8	14	13	7	3	1	
201	1	4	8	11	15	9	5		
202	3	3	5	8	12	13	4	5	
203		1	4	7	11	10	7	7	6
204	12	4	13	9	6	3	3	3	
205	2	7	9	10	14	8	3		
206	6	9	5	15	10	4	3	1	
207	3	3	8	7	13	11	4	2	2
208	4	13	9	10	9	4	3		1
209		3	6	8	17	13	4	2	
210	4	10	10	14	9	5		1	
211		11	9	15	13	4	1		
212	3	5	7	10	8	9	5	5	1
213	1	3	4	10	9	8	6	7	5
214	2	4	7	7	4	7	8	7	7
215	3	1	6	14	10	9	4	5	1
216	1	4	2	4	11	14	9	6	2
217	3	3	8	9	8	13	6	2	1
218		1	4	11	16	5	10	6	
219	4	7	13	9	15	2	3		
220		4	5	4	10	17	6	4	3
221	4	12	8	7	6	8	3	4	1
222	2	2	8	17	15	5	3	1	
223	4	4	4	5	6	10	5	8	7
224	2	8	9	15	14	1	1	3	
225		3	10	17	14	7		1	1
226	1	2	5	10	18	9	4	3	1
227	1	1	8	11	10	10	6	4	2
228			3	5	3	12	6	7	17
229	4	13	13	11	11	1			
230		4	2	6	8	13	10	7	3
231	1		1	4	11	8	10	13	5
232	2	3	7	10	15	5	4	5	2
233		3	4	12	20	10	2	1	1
234		1	3	7	13	14	3	11	1
235		1	8	12	10	9	7	3	3
236	10	10	10	13	7	1	1	1	
237	2	2	6	7	17	6	7	3	3
238	1	2	5	4	9	9	14	6	3
240			1	7	10	12	9	8	6

Student	Grade								
	1	2	3	4	5	6	7	8	9
241	1		2	6	8	6	12	12	6
242	1	1	3	6	22	11	6	2	1
243	13	13	8	12	4	2	1		
244		2	12	11	11	9	4	1	3
245	20	11	11	3		5		1	2
246	4	6	6	13	5	7	6	3	3
247	3	6	14	13	5	8	1	2	1
248	1		4	3	13	10	11	6	5
249		3	8	5	12	9	10	3	3
250	3	5	12	10	10	5	4	3	1
251	2	8	4	7	12	8	7	3	2
252	1	3	8	7	10	11	7	5	1
253	4	4	11	16	13	5			
254	1	2	9	12	8	3	8	5	5
255	2	2	8	9	11	8	3	6	4
256		1	4	15	13	7	9	4	
257			2	7	8	12	8	3	13
258			3	4	12	10	13	7	4
259		1		5	9	8	11	15	4
260	2	2	8	13	12	10	4	1	1
261	1	2	7	11	10	12	7	1	2
262	6	5	10	13	11	2	3	1	
263	4	7	13	8	9	4	5	1	2
264	8	8	10	10	9	7		1	
265	2	2	6	12	15	8	5	2	1
266	4	9	13	9	9	5	4		
267	2	8	10	14	13	3	2	1	
268	1	2	3	3	10	11	14	5	4
269			6	10	20	8	6	3	
270	6	7	21	4	11	3	1		
271	4	6	6	12	9	7	6	3	
272		3	11	16	12	4	6	1	
273		3	7	15	17	7	3	1	
274	44	5	2			1	1		
275			4	6	6	6	10	10	11
276	13	20	7	9	1	2	1		
277		1	3	5	10	11	11	7	5
279	1		8	15	11	7	7	2	2
280	2	3	9	9	10	13	3	3	1
281	4	6	14	12	11	3		2	1
282			3	6	11	13	10	7	3
284		5	8	17	13	5	3	2	
285	1		2	10	20	7	10	2	1
286	13	14	17	4	1	2	1	1	
287	1	4	4	7	14	12	2	8	1
288		1	6	13	22	6	4	1	
289	3	6	14	17	10	1	2		
290	3	7	13	13	9	3	5		

Student	Grade								
	1	2	3	4	5	6	7	8	9
291	3	22	14	6	5	1	2		
292	3	3	7	9	9	10	7	4	1
293	12	7	14	9	6	2	2	1	
294	1	1	1	4	6	11	5	12	12
295	1		3	4	8	7	10	12	8
296	1	1	4	9	13	18	6		1
297	6	8	15	12	4	5	3		
298	21	13	9	7	1	1	1		
299	3	2	5	3	10	7	9	10	4
300		2	7	11	17	8	5	2	1

Appendix H

Sample Score-Sheet

Reader ____	NEUTRAL	NOT BOTH. BY	POOR	GOOD	UNCLAS	LENGT OF COMME
QUAL. IDEA						
Amt. Id.						
Clar. Id.						
Orig. Id.						
Agree. Id.						
Coher. Id.						
Devel. Id.						
Persuas.						
Analytical						
Logic						
Pr. Opp. Id.						
Qual. Info.						
Qual. Illus.						
Amt. Illus.						
Generality						
Relevance						
Sincerity						
Interest						
STYLE						
Maturity						
Clar. Exp.						
Effect. Exp.						
Orig. Exp.						
Humor						
Variety						
Beginning						
Middle						
End						
GEN. ORGAN.						
Transitions						
No. Para.						
Conciseness						
Brief						
Long						
GEN. PARA.						
Brief Para.						
Long Para.						
GEN. SENT. STR.						
Variety S.						
Type S.						
Short S.						
Long S.						
Phrasing						
MECHANICS						
Grammar						
Punctuation						
VERBAL FACIL.						
Vagueness W.						
Ornateness						
Jarg., Sl., Col.						
Clichés						
Wordiness						
W. Choice						
W. Form						
Spelling						

Appendix I

Review of John B. Carroll, "Vectors of Prose Style" (1960)

It seems to me that Carroll shows what adjectives go together with other adjectives in describing a sample of objects in a given class, and that these are not necessarily important discriminants of prose styles.

The six "vectors" (related clusters of adjectives) that he found were represented as the following six scales:

Good	Personal	Ornamented	Abstract	Serious	Characterizing ("he is....")
Bad	Impersonal	Plain	Concrete	Humorous	Narrating ("he does....")

It should be obvious that the passages rated "good" might be either personal or impersonal, ornamented or plain, abstract or concrete, serious or humorous, characterizing or narrating. Similarly, the passages rated "personal" might be either good or bad, ornamented or plain, etc. These scales are logically "orthogonal" to one another: that is, standing on one has no necessary connection with standing on any other.

But consider the scales that had "loadings" on the "good-bad" dimension: pleasant-unpleasant, strong-weak, interesting-boring, graceful-awkward, varied-monotonous, clear-hazy, etc. Obviously no group of judges will say that certain passages are "good" but at the same time unpleasant, weak, boring, awkward, monotonous, hazy, etc. It might be possible to find one passage that exhibited such a violent reversal of form but not 150 representative passages rated by eight judges.

The dimensions that one gets by this procedure depend on the collection of adjectives one starts with. Carroll starts with just 29 pairs of adjectives without claiming that they represent all the different kinds of adjectives used by critics. A different initial collection might yield the following dimensions, each "orthogonal" (or nearly so) to one another and to all of the dimensions listed above:

Scientific	Neutral	Formal	Archaic	Outspoken	Bitter
Literary	Persuasive	Informal	Modern	Restrained	Hopeful

These are the first six additional dimensions that come to mind that might turn out to be independent. I can see no limit to the number of independent dimensions that might be found if one started with a sufficiently large and varied collection of adjectives. There would be no way to find out which dimensions are more important than others for any given purpose; the procedure can only show that they are independent. As an incidental outcome, it also shows which qualities were most reliably rated, but these are not necessarily the most important.

Carroll recognizes this objection in the sentence, "Even though somewhat comparable results might be obtained by asking raters simply to judge adjectives for similarity of meaning in the abstract, the dimensions of meaning cannot exist without some support in the real world to which the adjectives presumably refer."

True; one has to judge the compatibility or incompatibility of adjectives as applied to objects in a given class. Within the class that Carroll has examined -- 150 samples of English prose -- it seems to me that he has found six clusters of adjectives that go together out of the 29 pairs he started with, but that these are not necessarily important discriminants of prose styles.

Paul B. Diederich, ETS

Appendix J

Glossary for English Teachers

The basic procedure used in this study was the correlation of each reader's grades with each other reader's grades and with scores on three College Board tests. A correlation is a decimal representing the closeness of agreement between two sets of grades or other measures. It is not a percent and should not be thought of as "a correlation of x percent" -- as though x percent of the students received the same grade from one reader as from the other. It is affected both by the number of students who received identical grades from the two readers and by how close the remainder came to getting the same grades.

Nearly all the correlations found in this study were extremely low. The median correlation between the grades of any two readers was .31; the median correlation between the grades of any two college English teachers was .41. By contrast, selected College Board readers who have been trained to judge papers on a particular assignment by the same standards can sometimes hit an average correlation of .70 with one another on that assignment.

Another reference point is the fact that the correlation between height and weight among adults of the same sex is usually about .55. Tall people tend to weight more than short people, but the correspondence is not very close, for some are tall and thin while others are short and fat. The median correlation of .31 between readers' grades in this study shows that they agreed with one another far less than correspondence between height and weight.

Factor analysis is a mathematical procedure that picks out the basic dimensions or tendencies in a table of correlations. The details of the factor analysis and tables pertaining to it appear as Appendix A. The large table of 1,540 different correlations used in the factor analysis appears as Appendix B. In our table of correlations, we found just five dimensions or "factors" for the readers, representing tendencies or "schools of thought" in grading the papers. These schools of thought are superimposed on a considerable amount of reader agreement and an even larger amount of error (random variation). For each reader there is a decimal called a "loading" on each of these factors. The higher the loading, the closer a reader comes to the central tendency represented by that factor. Readers who have high loadings on a certain factor agree with one another better than with readers who have low loadings on that factor or high loadings on other factors. Readers who have nothing but low loadings fail to agree well with anyone or simply have a great deal of error in their grades. That is, they might not even agree with themselves if they were to do the job over again at a later time.

One type of factor analysis first blocks out the amount of agreement that is common to all readers and calls it a "general factor." It then shows the remaining differences between groups of readers who agree with one another better than with members of other groups. These remaining tendencies that differ from the general factor and from one another are called "group factors." Frequently they are difficult to interpret. Although we found the general factor of reader agreement to be large, we adopted a method designed to show most clearly the differences among "schools of thought" in grading the papers by finding the group factors without first subtracting what they have in common. The reported correlations between factors (Table A-3, page A-9, in Appendix A) show the amount of agreement between these group factors or "schools of thought."